EXPERIMENTAL DESIGN FOR SIMULATION

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Management Science

at the

University of Canterbury

by

Twan A.J. Vollebregt

University of Canterbury

February 1996

TABLE OF CONTENTS

ABSTRACT

Classical experimental design methods have gained widespread acceptance in the simulation literature. The simulation experimental design literature concentrates almost exclusively on factorial, fractional factorial, and composite simplex designs, which can be significantly more efficient than random or ad-hoc methods. However, there are several substantial differences between the classical (statistical) and simulation contexts that have received little attention. Most importantly, the design literature concentrates on obtaining maximum information from a set number of experiments, while in simulation we often wish to obtain a given amount of information at minimum cost. Also, classical designs and design methods generally assume constant variance and constant cost-per-experiment, while this is generally not the case in simulation. Hence classical designs are often not suitable for the simulation context. In addition, there are few rules to guide the experimenter in choosing an appropriate design, leading to quite arbitrary selection procedures. Thus although computer simulation is the ideal environment for which to develop experimental design software, the limitations of classical design methods mean that such software would do little more than perform routine tasks.

In this thesis we discuss the main differences between the classical and simulation contexts, and propose and develop an alternative design approach that is often more suitable for the simulation context. The design for our approach is found by solving an optimisation problem, and includes an element of sequentiality. In conjunction with a proposed solution heuristic, our approach is easily incorporated into experimental design software that requires little input from the experimenter. A number of examples and a Monte Carlo study are presented to illustrate the properties of our approach. We also discuss sequential experimental design, and list a number of research issues.

ACKNOWLEDGEMENTS

I would like to thank the following people:

- Don McNickle and Krzysztof Pawlikowski, for suggesting the general topic, providing supervision, and reading numerous drafts,

- John Deely and John George, for their willingness to discuss statistical and optimisation aspects of my thesis respectively,

- Ralph Disney, for providing useful comments on a draft of my thesis,

- The University Grants Committee, for providing financial assistance.

INTRODUCTION

The research reported in this thesis is concerned with experimental design, in the context of stochastic simulation. Current experimental design methods for simulation are critically examined, and existing alternative methods investigated. A new approach to experimental design is then proposed, and its properties discussed and illustrated through a number of examples. This thesis finished with a discussion of an area for future research, namely sequential experimental design.

The focus in the initial stages of the research was on the more general area of Response Surface Methodology (RSM). This methodology consists mainly of experimental design theory, parameter estimation methods, and function optimisation methods. Chapter 1 provides a review of the literature in this area. Many different design, estimation, and optimisation methods can be used as part of RSM. The initial research objective was to determine which of the many different methods suggested in the literature were suitable for practical implementation into experimental design software. Such an implementation would ideally allow practitioners to use the methodology to answer quantitative questions about a simulation model, without requiring (i) full knowledge of the theory behind the methodology, and / or (ii) input data that the practitioner is unlikely to have.

The component of Response Surface Methodology which appears to have received the most attention in both the simulation and general statistical literature is experimental design. However, a closer investigation into experimental design theory revealed that there were no obvious candidate methods in the design literature that are suitable for practical implementation in software. In addition, many design methods that were originally developed for an agricultural context appear to have been applied in the simulation context with little consideration as

to the difference between the contexts, and the implications of a number of classical assumptions.

As a result, the focus of this research shifted away from the general Response Surface Methodology area, and towards experimental design theory. The aim of the research then became the development of experimental design theory specifically for the simulation context. A major emphasis of the research was to develop a design method that can easily be coded up in software, and requires a minimum of user input.

In Chapter 2 we critically examine the literature on experimental design for simulation, and conclude that there are a number of concerns with the application of classical design methods to simulation. We then propose a new approach to experimental design in Chapter 3. This approach is similar to the classical optimal design approach, in that the design is found by solving an optimisation problem (design problem). However the focus of our approach is quite different from the classical approach, and we introduce an element of sequentiality. In Chapter 4 we consider the elements of the design problem, and how they can be selected. We also investigate solution methods, and develop a solution heuristic. Some properties of our approach are illustrated in Chapter 5. Finally, in Chapter 6 we discuss fully-sequential experimental design, and the research issues that must be resolved before sequential design methods can be developed.

In order to limit the scope of this thesis, the research reported here is restricted to the investigation of a particular scenario. We assume that a verified and validated simulation model is the subject of investigation. A single response is obtained from each simulation experiment, and this response is assumed to be an independent random variable. As in most of the design literature, the form of the response surface model is assumed to be known, including the difficult question of which factors should be included.

We discuss only terminating simulation (a subset of finite horizon simulation) and steady-state simulation (a subset of infinite horizon simulation). Similarly, we do not discuss simulation techniques such as run-length control and variance estimation methods in detail. Rather, we aim to develop an experimental

design framework that is largely independent of specific simulation situations and techniques.

CHAPTER 1: BACKGROUND

1.1. Introduction to Response Surface Methodology

The term Response Surface Methodology (RSM) is commonly used to refer to a collection of mathematical and statistical tools, which can be applied to obtain and study a 'response surface'. Some examples of those tools are the theory of experimental design, multiple linear regression, canonical analysis, and numerous function optimisation methods.

In brief, RSM involves estimating and analysing the form and parameters of a function, relating a response (yield variable) to one or more factors (stimulus variables) that are assumed to influence the response, and possibly determining an optimal factor combination. When RSM is applied to simulation, the resulting *analytical response model* is a model of the *simulation model*, which in turn is a model of a *real-life system*. As a result, the analytical model is often called a 'metamodel' in the simulation context (Kleijnen (1975)). Figure 1.1. shows the relationship between the real-life system, the simulation model and the metamodel.

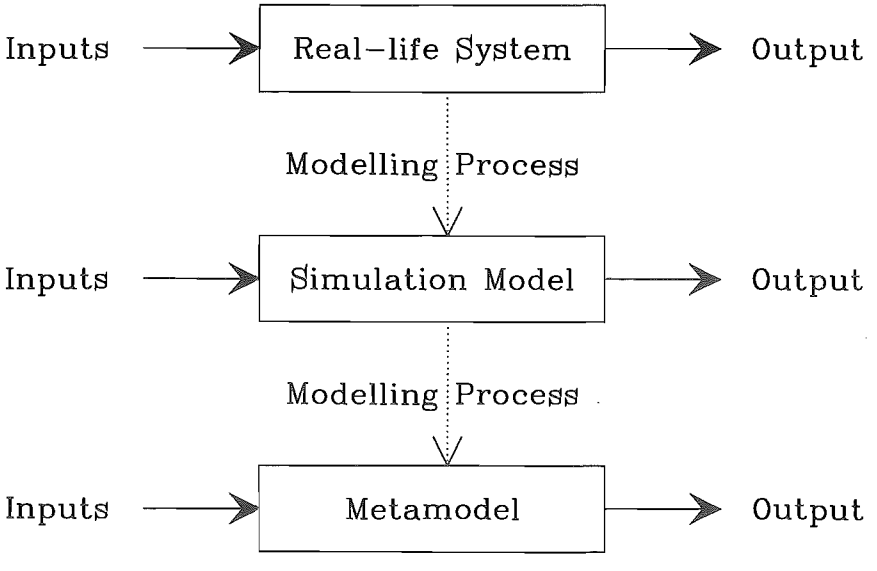


Figure 1.1. The modelling process

The main motivation behind finding an algebraic model for the system (or simulation model) under study, is that it is often costly to determine the response of the system for given factor settings. For example, in agricultural experiments it may take months before the yield of a crop can be measured. Similarly, runs of a complex simulation model may also take a substantial amount of time and computer resources. However an algebraic model, found using the results from a small number of experiments with the system, may be used to quickly and cheaply indicate the answer to any future questions about the system under study.

Applications of RSM arise in any field where a relationship between the factors and response of a system, and/or optimum factor setting, is to be determined by a process of experimentation. The primary area of application has traditionally been in the agricultural and chemical sciences, although frequent mention of RSM is also made in the computer simulation literature.

Probably the most influential paper on the research into RSM was Box and Wilson (1951), which brought together the various components of RSM into a methodology. Many of the fundamental ideas had been used and discussed much earlier. Mead and Pike (1975) provide an extensive account of the historical development of RSM through to 1975.

There have been many literature reviews done that are relevant to RSM. General reviews of RSM have been done by Hill and Hunter (1966), Mead and Pike (1975), and Myers, Khuri and Carter (1989). Reviews of simulation optimisation have been done by Farrell (1977), Meketon (1987), Jacobson and Schruben (1989), Safizadeh (1990) and Fu (1994). Steinberg and Hunter (1984) provide a general review of experimental design, while Donohue (1994) provides a review of experimental design for simulation.

The remainder of this chapter is a brief overview of the RSM literature, to provide background and motivation for the focus of this thesis: Experimental design for simulation. Special emphasis is placed on literature reporting on empirical studies and software implementation.

Sections 1.2. through to 1.5. correspond with the major stages of an RSM study: Selection of the model and parameter estimation methods, experimental

design, model analysis, and optimisation. A flow-chart relating these stages is shown in Figure 1.2.

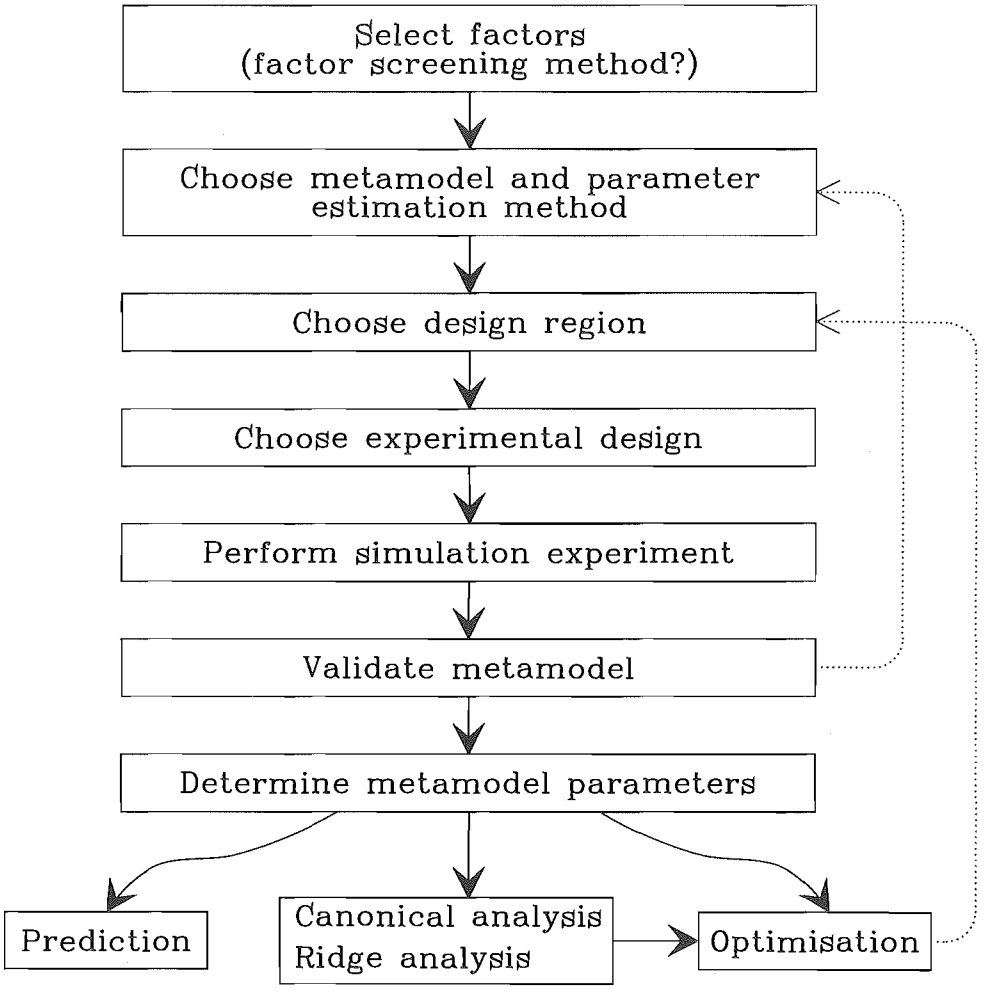


Figure 1.2. Some possible stages of an RSM study

An extensive literature exists on many of the tools considered to be a part of RSM. For example, the literature on experimental design is a substantial part of the statistic literature. However, the literature considered in this chapter is generally restricted to literature that places those tools in the context of RSM.

1.2. Selection of the Metamodel and Parameter Estimation Method

Let the relationship between the response η and the vector of factors ω of the simulation model be represented by the function Φ_1:

$$\eta = \Phi_1(\omega). \tag{1.1}$$

The vector of factors ω consists of all of the inputs to the simulation model. This includes any random number stream seeds that are used in stochastic simulation models.

The function Φ_1 takes the form of a simulation model. Responses from such a model can take a substantial amount of time and computing resources to obtain, and by themselves provide little insight into the relationship between η and ω. On the other hand, an analytical metamodel would allow such insight to be gained. Let the analytical form of the metamodel be represented by the function Φ_2:

$$y = \Phi_2(\mathbf{x}, \beta), \tag{1.2}$$

where y is the response, $\mathbf{x}$ is a vector of metamodel factors, and β is a vector of metamodel parameters. The metamodel factors may be simulation model factors, or they may be functions of simulation model factors. One of the main aims of RSM is to determine estimates of the metamodel parameters.

In general the experimenter may (i) not know the correct form of Φ_2, and (ii) not know exactly which simulation model factors (and functions of these factors) should be included in (1.2). Hence the metamodel generally only provides an approximation to the simulation model. Also, it is important to consider the range of (metamodel) factor values for which the metamodel is assumed to be valid. The literature deals with two such ranges, referred to as the *region of operability* and the *region of interest*. The region of operability $\mathcal{R}_o$ is defined as the region of factor values in which experiments are able to be performed; in simulation this may be the region in which the simulation model is considered to be a valid model of the real system. For some situations this region

may be infinitely large, while in other situations it may be limited by constraints related to the factors of the simulation model. On the other hand, the region of interest (also known as design region) $\mathcal{X}$ is defined as a sub-region within $\mathcal{R}_o$, which the experimenter is currently exploring. This is the region with which RSM is mainly concerned. Except for the IMSE method discussed later in this section, it is usually assumed that (1.2) is a sufficiently accurate approximation to (1.1) within the region of interest, so that the effect of bias in the selection of (1.2) is negligible.

The form of Φ_1 most commonly used in RSM is the linear metamodel

$$y(\mathbf{x}_i) = \mathbf{f}^T(\mathbf{x}_i)\beta + \varepsilon_i, \tag{1.3}$$

where i denotes a particular experiment, $y(\mathbf{x}_i)$ is the measured response, $\mathbf{x}_i = [x_{1i}, x_{2i}, ..., x_{mi}]^T \in \mathcal{X}$ is an (m x 1) vector of factor settings (also known as design points), $\mathbf{f}(\cdot) = [f_1(\cdot), f_2(\cdot), ..., f_p(\cdot)]^T$ is a *known* continuous mapping of the design region $\mathcal{X}$ into $\Re^p$, $\beta = [\beta_1, \beta_2, ..., \beta_p]^T$ is a (p x 1) vector of parameters, and ε_i is the experimental error. Typically the errors from individual experiments are assumed to be independent, with $E[\varepsilon_i] = 0$ and $var(\varepsilon_i) = \sigma^2$.

Due to its simplicity, a commonly used form of $\mathbf{f}(\mathbf{x})$ is a simple polynomial of low order, so that

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 & x_1 & \cdots & x_m & x_1x_2 & \cdots & x_m^k \end{bmatrix},$$

typically with k = 1 (first order) or k = 2 (second order). The fact that a k^{th} order polynomial is the k^{th} order Taylor series expansion of the true model in the region of interest, provides some theoretical justification for this type of model.

The main advantage of using a linear metamodel is that the method of Ordinary Least Squares (OLS) can be used to estimate the parameters of such a model, and find goodness-of-fit measures. OLS is relatively easy to apply and well known. Let N be the total number of responses observed. Then from any standard regression text (e.g. Draper and Smith (1981)), the OLS estimator of the model parameters is given by

$$\hat{\beta} = \left(\sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1} \sum_{i=1}^{N} y(x_i)\mathbf{f}(x_i) .$$

Provided (1.3) correctly models the relationship between the factors and the response, and $\text{var}(\varepsilon_i)$ is constant, then this estimator is the best linear unbiased estimator (BLUE). Note that the matrix in brackets above is more often written as F^TF, where F is a matrix whose N rows are $\mathbf{f}^T(x_i)$. However, the summation form above more clearly shows the dependence of $\hat{\beta}$ on $\mathbf{f}(x_i)$ and N.

Assuming independence between the N responses $\{y(x_1), y(x_2), \ldots, y(x_N)\}$, the covariance matrix of the above estimator of β is given by

$$\text{Cov}(\hat{\beta}) = \sigma^2 \left(\sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}(x_i)^T \right)^{-1} .$$

Let s^2 be an unbiased estimator of σ^2. An estimate of the variance of the fitted response at any point $\mathbf{x}'$ is then given by

$$\text{Var}(\hat{y}(\mathbf{x}')) = s^2 \mathbf{f}^T(\mathbf{x}') \left(\sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}(x_i)^T \right)^{-1} \mathbf{f}(\mathbf{x}') .$$

Assuming that the N responses are normally distributed, a $100(1-\alpha)\%$ confidence interval for the fitted response is given by

$$\hat{y}(\mathbf{x}') \pm t_{N-1,1-\alpha/2} \sqrt{\text{Var}(\hat{y}(\mathbf{x}'))} .$$

Barton (1993) provides a survey of alternative methods for fitting a model to response data, such as Taguchi methods, kernel smoothing, and methods based on splines. However as noted by Barton, the method of least squares is widely known, and involves straightforward calculations. Unlike many other methods it is able to provide confidence intervals and other measures of goodness-of-fit, if the responses can be assumed to be (approximately) normally distributed.

However, it must be noted that there are two sources of error that will cause a discrepancy between the actual response of the simulation model and the fitted

response: bias error (lack of fit) and variance error (experimental variation). Generally the assumption is made that the latter source of error is substantially larger than the former. Karson, Manson and Hader (1969) present an alternative to least squares for situations where bias error predominates. They note that the method of least squares assumes that the metamodel is able to correctly model the relationship between the mean simulation response and the metamodel factors, and hence aims to minimise variance error. They propose a minimum bias estimator, which uses any additional flexibility to satisfy other criteria. However, this approach depends on the ability to specify the correct form of the metamodel, which is presumably unknown in situations where there is sufficient bias error to consider this approach.

To achieve a better fit, without loss of the practicality of models linear in their parameters, transformations can be applied to non-linear models. One example is the non-linear model $y = e^{\beta x}$, which can be transformed to the linear model $\ln(y) = \beta x$. Box and Cox (1964) discuss transformations of the dependent variable that lead to or preserve the assumptions of $E[e_i^2] = \sigma^2$ (where e_i is the observed value of ε_i), normality of e_i, and independence between the e_i. Lindsey (1972) uses maximum likelihood estimators for the parameters of a transformation of both the response and factors. Lastly, Box and Draper (1982) show how the transformation parameters for a power transformation of the factors can be estimated.

Most of the RSM literature considers RSM to be based around polynomial metamodels that are linear in their parameters. However, some authors include non-linear models under the RSM umbrella, and there is certainly an extensive literature of this topic in the biological sciences (see Mead and Pike (1975)). In that field there are often good theoretical reasons for assuming a particular non-linear relationship. However, the selection of a non-linear model leads to a significantly more complex estimation procedure (Rawlings (1988)).

1.3. Experimental Design

A classical experimental design E^c is most commonly defined to be the collection of pairs

$$(x_1, p_1), (x_2, p_2), \ldots, (x_r, p_r),$$

where r is the number of distinct design points x_i (as factor settings are labelled in the design literature) at which the *proportion* $p_i > 0$ of experiments is to be performed. The design points are usually considered to lie within the region $\mathcal{X}$, which is known as the design region. Alternatively, the pairs (x_i, n_i) can be considered, where n_i is the *number* of experiments to be performed at x_i. The relation between the two definitions is $p_i = n_i / N$, where $N = \sum n_i$.

Experimental design has been an important part of the RSM methodology, and certainly a look at the literature reveals that the majority of published papers relating to RSM focus on experimental design issues. In the general statistical literature, experimental design has also received much attention. The main reason for this is that the choice of the design can have a large impact on the quality and usefulness of the data obtained by the experiments. This can be seen by considering the covariance matrix of the estimated parameters of (1.3),

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}) &= \sigma^2 \left(\sum_{i=1}^{N} \mathbf{f}(x_i)\mathbf{f}(x_i)^T \right)^{-1} \\ &= \left(\sum_{i=1}^{r} \frac{n_i}{\sigma^2} \mathbf{f}(x_i)\mathbf{f}(x_i)^T \right)^{-1} \\ &= M^{-1}. \end{aligned} \tag{1.4}$$

For later reference the matrix between the large brackets on the second line of (1.4), also known as the Fisher information matrix, has been defined as M. Equation (1.4) shows that the variability of the parameter estimates depends directly on the choice of experimental design. Through careful manipulation of the position of the design points and the number of experiments allocated to them, we can gain considerable efficiency.

It has been common in the literature to begin by standardising the factors, so that the effect of scale is removed from any analysis of the design (e.g. Box and Draper (1959), Smith (1973b), Safizadeh (1990)). In addition, standardisation allows standard designs to be tabulated. The standardised factor z_i is related to the actual factor x_i by:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i},$$

where $\bar{x}_i$ is the centre of the design region along the x_i axis, and S_i a scaling constant. In some papers, S_i is defined as

$$S_i = \left[\sum_j \frac{(x_{i,j} - \bar{x}_i)^2}{N} \right]^{1/2}$$

(e.g. Safizadeh (1990)), while in other papers S_i is set so as to ensure that $-1 \leq z_{ij} \leq 1 \ \forall i,j$ (e.g. Smith (1973b)). Standardised factors particularly lend themselves to the formation of orthogonal designs, for which $f(\mathbf{x}_i)^T f(\mathbf{x}_j) = 0 \ \forall i \neq j$. Such designs have the desirable property that the estimated parameters are uncorrelated, since the Fisher information matrix M for such designs is a diagonal matrix. However, it appears that standardisation has lost some of its popularity (Safizadeh (1990)).

However, apart from efficiency and orthogonality there are a number of other desirable properties that a design can have. According to Box and Draper (1975), other properties of a good design are that it should:

i. generate a satisfactory distribution of information throughout the region of interest,
ii. ensure that the fitted response be as close as possible to the true response,
iii. give good detectability of lack of fit,
iv. allow transformations to be estimated,
v. allow experiments to be performed in blocks,
vi. allow designs of increasing order to be built up sequentially,

vii. provide an internal estimate of error,
viii. be insensitive to wild observations and to violations of the usual normal theory assumptions,
ix. require a minimum number of experimental points,
x. provide simple data patterns that allow ready visual appreciation,
xi. ensure simplicity of calculation
xii. behave well when errors occur in the setting of the predictor variables,
xiii. not require an impractically large number of predictor variable levels,
xiv. provide a check on the 'constancy of variance' assumptions.

It is clear from this list that there are many design properties for the experimenter to consider when selecting a design for a particular situation. Depending on the situation, some properties may be more important than others, and different amounts of prior information on the form of the metamodel and the response data may be available. As such, it is not surprising that there have been many different approaches to experimental design.

To most clearly explain the differences between the various experimental design approaches, we will categorise them as follows: (a) design property criteria, (b) variance-optimal design criteria, and (c) J-optimal design criteria.

(a) *Design property criteria* are concerned with one or more *properties* of the design. Typically the objective is to make sure the design exhibits the desired properties, rather than minimising the deviation from those properties. Some of the more common design properties found in the literature are:

Orthogonal designs are termed so because $f(\mathbf{x}_i)^T f(\mathbf{x}_j) = 0 \ \forall i \neq j$. This has the advantage that the estimates of the parameters of the fitted model will be uncorrelated provided the variance of the responses is constant (Khuri and Cornell (1987)). However orthogonality is generally restricted to 'first-order' designs (designs for first-order polynomial metamodels). Both first-order factorial and fractional factorial designs (see below) are orthogonal.

Saturated designs have the property that the number of design points is equal to the number of parameters in the fitted model (Box and Draper (1987)). Since at least as many design points are required as there are parameters in the model in order to determine the parameters, saturated designs have the minimum number of design points.

Rotatable designs were developed for fitting second and higher order polynomial models, and can be constructed by combining the vertices of regular geometric figures plus centre points (Hunter and Naylor (1970)). Their construction guarantees that the variance of the fitted response at any point depends only on the distance from the centre of the design, not the direction. This can be shown using the function $V(\mathbf{x}) = N \, var(y(\mathbf{x})) / \sigma^2$, first used by Box and Hunter (1957). The condition for rotatability is that $V(\mathbf{x})$ is constant along a (hyper-) sphere in the factor space with origin at the design centre. Rotatability is important when the orientation of the actual model relative to the axes is not known, so that the design does not need to be rotated for better estimation precision.

Uniform precision designs are a subset of the rotatable designs. The extra condition put on them is that the value of $V(\mathbf{x})$ at the design centre is equal to that at the design points (Myers (1971)). This leads to a more uniform 'precision' over the design region.

In the literature, several standard design types have emerged by applying one or more of the above criteria:

Factorial designs have $L_1 \times L_2 \times \ldots \times L_p$ design points, where L_p is the number of levels of factor p. Factorial designs assume that the design region is a (hyper-) cube, and for $L = 2$ they require that the design points are equally spaced along the vertices of the design region. It allows calculation of the main effect of a factor, such as x_1 and x_1^2, and interaction effects between the factors, such as x_1x_2, up to p-way interactions (Mead (1988)).

Fractional factorial designs attempt to reduce the large number of experiments needed by a factorial design. This is done by using only part of a factorial design, and sacrificing the ability to estimate higher-order interactions. Depending on which part of the factorial design is chosen, this may lead to an inability to distinguish some lower order effects from higher order effects, as there may be fewer design points than parameters in the model to be fitted. Steinberg and Hunter (1984) and Hunter and Naylor (1970) provide numerous references on both factorial and fraction factorial designs, and Raktoe, Hedayat and Federer (1981) provide references to research on factorial design construction methods.

Simplex designs are first-order saturated orthogonal designs consisting of p design points, which are located in such a way that the distance between any two points is equal (Box and Draper 1987)).

Central composite designs are probably the most commonly used second (and higher) order designs (Myers, Khuri and Carter (1989)). They are constructed using standardised variables, by assembling (a) a 2^p factorial (or fraction thereof) set at a distance α, (b) 2p axial points set at a distance ß, and (c) p_0 centre points (Biles and Swain (1979)). The flexibility of central composite designs comes from the ability to influence for example rotatability and model mis-specification robustness through the choice of α (Myers, Khuri and Carter (1989)), and orthogonality through the choice of p_0 (Biles and Swain (1979)). For example, the condition for rotatability is $ß/\alpha = 2^{p/4}$.

Meeker, Hahn and Feder (1975) report on the development of a computer program for the evaluation and comparison of designs according to standard properties.

(b) *Variance-optimal design criteria* can be used to select the design that minimises the variance of some function (usually of the model parameters) known as the design criterion. This is in contrast to design property criteria, which focus on modification of a design so that it satisfies certain design properties, rather than selection of the design itself. Since the publication of a

paper by Kiefer and Wolfowitz (1959) there has been a large amount of research published in this area, commonly known as optimal design theory. A good general review is Steinberg and Hunter (1984), who provide references to several more reviews on this topic. However, in the simulation literature Cheng and Kleijnen (1995) appears to be the only paper that has investigated optimal design in a simulation context.

The procedure for determining an 'optimal design' is to specify the form of the model to be fitted, determine a suitable design criterion, and apply an algorithm to find the design that is optimal for the criterion. Central to any design criterion is the Fisher information matrix M defined in (1.4). Some of the more common optimality criteria found in the literature are:

- D-optimality: Minimise the determinant of M^{-1}
- A-optimality: Minimise the trace of M^{-1}
- E-optimality: Minimise the maximum eigenvalue of M^{-1}
- G-optimality: Minimise $\max\left(\mathrm{var}\left(\hat{y}(\mathbf{x})\right)\right)$ taken over all $\mathbf{x}$.

(Steinberg and Hunter (1984)). Note that there is a general equivalence theorem linking D- and G-optimality (Kiefer and Wolfowitz (1959)). Due to the labels, this category of design criteria is often referred to as alphabetic optimality. Atkinson (1982) discusses developments in this area during 1975-80.

However, finding a design that conforms to one of the alphabetic optimality criteria is not easy. One approach is algebraic solution, but this is generally reasonably complex. Another early approach was to use mathematical programming techniques (e.g. see Atkinson (1969), Box and Draper (1971)). Due to the problem of dimensionality, if the total number of experiments N and the number of factors p is large, obtaining a solution in this way can be time consuming. A number of heuristic procedures have also been considered. Most of these are exchange algorithms, which start with an N point non-singular design, and iteratively add and delete points to improve the criteria. The criterion is almost always taken to be D-optimality, or a close variant. Cook and Nachtsheim (1980) discuss seven algorithms for finding exact D-optimal designs, and compare their performance, while Welch (1982) presents a branch and bound

method for finding exact D-optimal designs, assuming a finite number of candidate sites.

A major difficulty with finding the optimal design is that the number of experiments allocated to a particular point must be integer. Kiefer and Wolfowitz (1959) overcame this problem by working with the *proportion* of the total experiments to be performed at each point, p_i. Approximate designs can then be found by making the design problem continuous, and later rounded to give integer designs. Provided N is reasonably large, the integer design is likely to be near optimal, although this is not guaranteed.

(c) *J-optimal design criteria*, unlike variance optimal criteria, are concerned with bias error as well as variance error. Much of the literature on experimental design has concentrated on variance error, whereas bias error has had little attention. However, in RSM applications in particular it is usually assumed that a simple polynomial model is used only as a local approximation. One of the effects of ignoring bias error is as follows:

> "Assuming a particular model, the variance of y(x) at some point x will normally decrease as the size of the design [distance between the design points] is increased, so that if variance error is treated as the only kind of discrepancy we are led to the conclusion that in order to obtain a good representation over [the region of interest] we ought to take as large a design as possible over the [region of operability]."

(Box and Draper (1959, p624-25)). This is clearly at odds with the fact that the ability of the fitted model to represent the true response will decrease as wider regions of interest are considered.

Probably the first paper to discuss the development of designs for protecting against model inadequacy was the paper by Box and Draper (1959) quoted from above. Their design criteria were that:

(a) the fitted model, a polynomial fitted by least squares, was to most closely represent the true response model (assumed to be a higher order polynomial) over the region of interest, and

(b) that there be a high chance of detecting model mis-specification, in particular if the true response model was a polynomial of higher order than the fitted model.

To meet requirement (a), the design is chosen so as to minimise the *expected mean square error* J, averaged over $\mathcal{X}$:

$$J = \frac{N}{\sigma^2 \int_{\mathcal{X}} dx} \int_{\mathcal{X}} E\left[\hat{y}(x) - y(x)\right]^2 dx,$$

where the error is made up of the sum of variance error and bias error:

$$\begin{aligned} J = & \quad \text{Variance error} \quad + \quad \text{Bias error} \\ = & \frac{N}{\sigma^2 \int_{\mathcal{X}} dx} \left[\int_{\mathcal{X}} \left(\hat{y}(x) - E\left[\hat{y}(x)\right]\right)^2 dx + \int_{\mathcal{X}} \left(E\left[\hat{y}(x)\right] - y(x)\right)^2 dx \right]. \end{aligned}$$

This criterion is usually called the integrated mean squared error criterion (IMSE). Khuri and Cornell (1987) point out that the process of averaging used may mask poor performance in a particular area, by considering the average error over the whole region $\mathcal{X}$.

The conditions that must be met for a design to be J-optimal are usually expressed in terms of restrictions on the moments of the design, derived algebraically (e.g. Box and Draper (1963)). Implicitly the moment conditions determine the number of repetitions to be made at each distinct design point. Note that a value for N is not needed to determine the moment conditions, but does need to be specified to completely define the design.

For the situation where a second degree polynomial is approximated by a first degree polynomial, Box and Draper (1959) come to the conclusion that the optimal design is nearly the same as the one for which bias error alone is minimised. Myers, Khuri and Carter (1989) and Donohue, Houck and Myers (1992) list further papers that support this conclusion, and both note that convincing arguments have been made for including bias in the selection of a design.

In the simulation literature, the primary application of the J-optimality idea has been to use it to modify standard designs, rather than to determine a design.

Because the conditions for a J-optimal design can be quite flexible, authors have taken one or more of the standard design definitions, such as central composite, factorial, etc, and made them J-optimal. For example, see Donohue, Houck and Myers (1992).

However, there are two main problems with the J-optimal approach. First, it requires the experimenter to specify the true response model that is generating the responses, or at least a model against which they want to be protected. Second, the J-optimal design will always depend on certain parameters of that model, which are (presumably) unknown. In essence, the problem of not knowing what the true response model is, is transferred to the problem of not knowing the parameters of its assumed form.

In previous sections it has been assumed that we are most interested in estimating either the parameters of the (meta)model to be fitted, or the mean response. However, in some situations the accurate estimation of the *slope* of the response surface in the region of interest may be more important. This will be the case when gradient techniques such as steepest ascent or ridge analysis are being used to move closer to an optimum or specified response level.

Atkinson (1970), Ott and Mendenhall (1972) and Murty and Studden (1972) all investigated designs for estimating the slope of a response surface. Atkinson uses Box and Draper's (1959) IMSE criterion to minimise the sum of bias and error variance of the directional derivative, averaged over all possible directions. Ott and Mendenhall consider the behaviour of the slope-variance of a second order, one-factor model. Murty and Studden also consider a one-factor model, and derive various slope-optimal designs using various characterisations.

Myers and Lahoda (1975) present the development of IMSE criteria for the estimation of a set of parametric functions. This is applied to the two situations where the partial derivative functions of a second (third) order model are estimated through a first (second) order model. In the first case, the optimal design properties include orthogonality and maximum spread, e.g. a 2^p factorial. In the second case, the design must at least be a second order rotatable design, and may be a composite design.

The concept of slope-rotatability for second order composite designs is presented by Hader and Park (1978). Slope rotatability implies that the variance of the slope is constant along a (hyper-) sphere in the factor space with origin at the design centre. This is done by setting α (see design property criteria section above), the error variance optimal values of which are higher than for ordinary rotatability. They also find that replications of the axial points, rather than the centre points, has advantages in this case. However, Hader and Park only considered rotatability of partial derivatives parallel to the factor axes. Park (1987) extends these results to the slope rotatability of central composite designs over all directions.

The experimental design literature has remained largely theoretical, usually being limited to the specification of new criteria, and methods for constructing designs. Little has been reported about how the different designs perform in various applications. Montgomery and Evans (1975) provide what appears to be the only such direct empirical research into which types of designs perform well. In their paper, 6 second-order designs were used to estimate the optimums of 6 different two-factor response surfaces. The experimental region is set up to cover the optimum, and only one experiment is used to determine the response surface, allowing calculation of the estimated optimum by canonical analysis (see next section). The designs used were:

- factorial
- orthogonal central composite
- uniform precision central composite
- minimum bias central composite
- orthogonal hexagonal
- uniform precision hexagonal.

Note that these are all rotatable designs. The conclusions were that orthogonal hexagonal design performed best overall, and also used relatively few design points. The minimum bias central composite design also performed well.

Another approach to the evaluation of designs is to compare them on the basis of various properties. For example, with standardised factors the variance of the fitted model's coefficients for various designs can be compared. This is the approach taken by Donohue, Houck and Myers (1992), who compare 4 types of standard property criteria designs that were made J-optimal. However, this approach assumes that the bias component has been correctly specified.

1.4. Model Analysis

Once the design has been determined, the experiments performed, and the metamodel fitted according to the method chosen, there are a number of methods for obtaining further information about the metamodel.

Canonical analysis can be applied to a second order polynomial metamodel, to provide further information about the type and location of a stationary point. The second order model

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{i=1}^{k} \sum_{j \geq i}^{k} b_{ij} x_i x_j \qquad (1.5)$$

can be re-expressed in matrix notation as

$$\hat{y} = b_0 + \mathbf{x}^T \mathbf{b} + \mathbf{x}^T \mathbf{B} \mathbf{x}, \qquad (1.6)$$

by defining

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{12} & b_{22} & \cdots & \frac{1}{2}b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}b_{1k} & \frac{1}{2}b_{2k} & \cdots & b_{kk} \end{bmatrix}$$

(Box and Draper (1987)). The eigenvalues (λ_i's) and eigenvectors (v_i's) of the matrix B are defined by $\mathbf{B}v_i = v_i\lambda_i$. Letting V be the matrix of orthonormal eigenvectors, we can write

$$\hat{y} = b_0 + \theta_1 x_1 + \cdots + \theta_k x_k + \lambda_1 x_1^2 + \cdots \lambda_k x_k^2,$$

where $\theta = V^T b$. This is the 'A' canonical form (Box and Draper (1987)) and it eliminates cross product terms by rotating the axes so that they are parallel to the principal axes of the system (the 'B' canonical form consists solely of an intercept and quadratic terms; it is obtained by also translating the origin of the axes to the stationary point).

By taking derivatives with respect to $\mathbf{x}$ in (1.5) the co-ordinates of the model's stationary point are found to be:

$$-2B\mathbf{x}_s = \mathbf{b}, \quad \text{or} \quad \mathbf{x}_s = -\tfrac{1}{2}B^{-1}\mathbf{b},$$

or in terms of the rotated axes:

$$x_i = \frac{-\theta_i}{2\lambda_i}.$$

The canonical form provides many clues as to the shape of the model. For example, the signs of the eigenvalues indicate the type of stationary point, e.g. max, min, saddle. The relative eigenvalue magnitudes provide an indication of the shape of the quadratic surface along the translated axes.

Ridge analysis is a procedure that can be used when canonical analysis has established that a ridge system (similar to a mountain ridge) is present. Such a system is characterised by zero or near-zero eigenvalues. In such a case, the optimum may still be far away from the current design (assuming there is a finite optimum), and any estimate of its location is likely to be inaccurate. In addition, note that a ridge does not necessarily follow a straight line. Hence we cannot simply sample at some points on such a line.

Ridge analysis was suggested by A. Hoerl and formally derived by Draper (1963). It starts by considering the second order response surface (1.5). We now restrict our attention to those points x_i which lie on a sphere of radius R, centred at the design origin. Somewhere on that sphere there will be a maximum of y, a

minimum of y, and possibly other stationary points. These can be found using the method of Lagrange multipliers, leading to the system

$$(B-\lambda I)x = -\tfrac{1}{2}b$$

(Box and Draper (1987)). This system of k+1 equations can then be solved for the k+1 unknowns. However a simpler method exists. This involves setting a value for λ, and using the above system to find x (and hence R). Depending on the value of λ that is used, we get various stationary points. In particular, setting λ to a value above the maximum eigenvalue of B will lead to an absolute maximum point. By substituting various values of λ (greater than the maximum eigenvalue) into the above system, we assemble a locus of points in k dimensions which represents a direction of steepest ascent (ridge analysis is hence the equivalent of steepest ascent for second order models). This information can be used to shift the design to a new region, where a better approximation of the model near the optimum can be made.

In general, there is very little in the literature on practical implementation of ridge analysis. Smith (1976) mentions that a ridge analysis module had been incorporated into his optimiser program. However, some authors recommend not using this technique, but without supplying a reason why (e.g. Box and Draper (1987)).

1.5. Optimisation

Most of the literature surrounding RSM has recognised that one of the primary objectives of the methodology is to determine optimum conditions. In fact, in their seminal paper Box and Wilson (1951) stated this as being the main reason for their study. However, looking at the literature not much research has been reported on finding good optimisation methods for RSM. Almost all methods used and described in papers are adaptations of reasonably well known non-linear optimisation methods. However, such methods were designed to be used in situations where evaluation of the objective function is not costly and the response is deterministic - certainly the opposite is true for stochastic models.

Also in the case of RSM it is likely to be beneficial to consider the interaction between the design of the experiment, the metamodel and the estimator used, and the optimisation method used. Again, little work appears to have been done in this area.

The literature on the optimisation phase of RSM can be classified according to the type of method used: Path search methods, and direct search methods.

Path search methods involve estimating a direction of movement, and a distance to move in that direction (Jacobson and Schruben (1989)). Essentially path search methods are variations on the method of steepest ascent, where the direction of movement is the gradient.

Steepest ascent has been the optimisation method most commonly associated with RSM. A typical procedure involves specifying an initial region of interest, and estimating the parameters of a first order polynomial response curve over this region. The vector of these estimated parameters $\left[\hat{\beta}_1 \quad \hat{\beta}_2 \quad \cdots \quad \hat{\beta}_p\right]^T$ is then used as the optimal direction, or vector of partial derivatives. A line search is then performed along this direction to obtain a step size δ^k. To do this, Biles (1974) suggests setting up an experiment along the gradient direction, fitting a polynomial model to the data, and calculating the step size by finding the maximum along this polynomial. The region of interest is then moved to the new point.

This is called Phase I. There are several criteria that can be used to decide when a first order model is no longer adequate, all concerning the detection of curvature:

(a) repeat Phase I until there is little improvement, signifying the possible presence of curvature,

(b) instead of a pure first order model we could fit a polynomial that includes an interaction term (such as x_1x_2) and terminate Phase I when it becomes clear that there is a significant interaction effect present (Safizadeh and Thornton (1984)),

(c) when n_1 observations with average Y_1 are collected at the points of a first order orthogonal design, and n_0 observations with average Y_0 at

the centre, and assuming that the true model is a quadratic, then Y_1-Y_0 is an unbiased estimator of the sum of the quadratic coefficients (Khuri and Cornell (1987)). An F-test can then be done on the significance of this difference.

(d) other general methods are those used to validate the regression model, such as the R^2 statistic, cross-validation etc (see Kleijnen (1992)).

Once significant curvature is detected, Phase II starts, which involves fitting a second order polynomial using a second order design. As the model is now a quadratic, canonical analysis can be used to bring the experiment to the optimum quickly, usually in one experiment. A final experimental design can be used together with canonical analysis to provide a final estimate of the optimum.

For Phase I, Brooks and Mickey (1961) investigated the optimum number of design points that would maximise the improvement per unit of effort, assuming that a first order model was a good approximation. They concluded that as few design points should be used as possible, such as those of a simplex design.

The major disadvantage of steepest ascent is that it usually requires many experiments, firstly because an accurate response surface needs to be estimated for gradient calculation, and secondly because it requires many iterations. The procedure is also not invariant to changes of scale in the factor units. On the other had, fitting the response surface will lead to considerable 'smoothing' of experimental error, and so lead to a more accurate gradient calculation.

Joshi, Sherali and Tew (1994) note that steepest ascent as used in RSM is a memoryless process, in that it does not use information from previous iterations to improve the search direction. Zigzagging may occur, leading to slow convergence. Also, the standard procedure after curvature has been detected is to fit only *one* second order model. They propose certain gradient deflection methods, augmented with restarting criteria, that improve the search direction. Computational results using standard test functions showed that one particular combination appears promising.

Direct search methods are the second main class of RSM optimisation methods. These methods progress through a sequence of points without using the

gradient (Biles and Swain (1979)). One such method, *pattern search methods*, uses some pattern in the observations to obtain an improved point (Jacobson and Schruben (1989)). However, they involve neither gradients, nor random methods. One advantage over path search methods is their ability to be applied to discrete problems.

The most common of the pattern search methods is the *Hook and Jeeves method* (Hook and Jeeves (1961)). The basic idea behind this method is that if a direction has produced a favourable change, then we should continue to move in that direction. Due to the random variation in the response, this pattern search method converges very slowly. However, Safizadeh (1990) suggests that it may be useful for quickly leading the investigation into a promising subregion.

Another pattern search method is the *Nelder and Mead sequential simplex method.* This is based on the unconstrained simplex method, which starts with a simplex of n+1 points (Avriel (1976)). The point $\mathbf{x}^h$ with the worst value of $y(\mathbf{x}^h)$ is then *reflected* through the centroid of the remaining points to point $\mathbf{x}^r$. This is the basic simplex method. The Nelder and Mead version, usually regarded as superior to the above version (Avriel (1976)), has additional steps. If $y(\mathbf{x}^r) > y(\mathbf{x}^i)$ for all $i \neq r$, then an *expansion* step is taken to further take advantage of the improvement in $y(\mathbf{x})$. On the other hand, if $\mathbf{x}^r$ becomes the new point $\mathbf{x}^h$, then a *contraction* step is taken to avoid introducing inferior points.

Some investigations of this method show that it is quite sensitive to the size and orientation of the initial simplex, and that it is very inefficient for problems with a large number of variables, e.g. $p > 10$ (Avriel (1976)). However, it works well relatively close to the optimum, and for very 'noisy' problems (Meketon (1987)). It is also easily modified to handle constraints (Biles and Swain (1979)).

There are four empirical studies of RSM optimisation methods that have appeared in the literature, all of which were done in the 1970's. Such studies are most easily performed by applying different methods to simulated output data, and this has been the procedure used in these studies.

The first, by Smith (1973a), was the most comprehensive. Seven search techniques were tested:

1. Random search over a specified region

2. Single factor search, in which a single factor is varied at a time until there is no more improvement
3. RSM variation I, which uses a 2^p factorial or 2^{p-k} fractional factorial design to determine an estimate of the gradient. Steepest ascent is then applied.
4. RSM variation II, identical to variation I except that it uses a simplex design.
5. Single factor with acceleration
6. RSM variation I with acceleration
7. RSM variation II with acceleration

The accelerated versions involve increasing the step size when several successive steps have resulted in an improvement. Each method was applied to known functions, and under different situations. Different situations were created by changing the number of controllable factors, the number of available computer runs, the presence or absence of local optima, the size of the random response error, the distance of the starting point from the optimum, the relative activity of the factors, and the presence or absence of factor interaction. Each method is able to use a fixed number of computer simulation runs.

The conclusions reached were that (a) the single factor method was completely dominated by other methods, (b) the accelerated techniques did not provide any improvement, (c) random search was feasible, surprisingly, provided the search region was defined carefully, and (d) the RSM variation I probably performed better than variation II.

In a paper concerned mainly with constrained multiple response optimisation, Biles and Swain (1979) test three methods on two problems. They found that a second order RSM method outperformed both a first order RSM method and a modified simplex method, in terms of number of trials needed.

Segreti et al (1979, 1981) examined the efficiency of steepest ascent, steepest ascent including the past three phases of data, and the simplex method. This was done in the context of a clinical trial situation, where a simulation was used to simulate the entry of patients to the trial, and their response to treatments. A simplex design was used in all cases, and the methods were tested on both planar and curvilinear response surfaces. Surprisingly, the steepest ascent version

that used past data performed best over all, even when applied to the curvilinear surface.

Computer simulation would seem the ideal application for a computer based stochastic response optimisation program. Probably the earliest attempt to design such a computer program was reported by Meier (1967). He reports on the development of a closed-loop program that uses a modified version of the sequential simplex method, described as a "general purpose optimisation program designed to be inserted into any simulation program" (Meier (1967, p33)). Emphasis was on treating the simulation program as a 'black box', which simply responds to the decision variable settings. Such a 'black box' treatment implies that no assumptions are made about the underlying model. Also, the linkage between the optimisation and simulation programs were kept to a minimum for portability.

Smith (1973b) also proposes the development of an optimiser, and discusses the design requirements it should have. Like Meier, he also suggests regarding the simulation model as a black box. The optimisation program should be independent from the simulation program for generality, and the user should not to have an extensive knowledge of either program: "In essence, the "Optimiser" would be obliviating the need for an "expert" ..." (Smith (1973b, p172)). In addition, the optimiser should be able to determine the best optimisation technique for a specific situation. Three design requirements are listed: Control of statistical variation, factor screening, and location and exploration of a region containing the optimum. The first two requirements were proposed so that the optimiser would have to deal with relatively few factors, and quite precise experimental data.

In a later paper, Smith (1976) reports that a modular optimum seeking program had been developed, which could be used for constrained or unconstrained optimisation. After an empirical comparison of various methods was made (see the section above), RSM with steepest ascent was chosen as the optimisation method. The search for an optimum was divided into:

1. First order design phase - this generates a 2^{p-k} fractional factorial design of minimal size

2. Steepest ascent phase - using the experimental data gained from (1) this phase monitors the simulation along the path of steepest ascent.
3. Factor screening phase - reduces the number of factors used in (4) by removing those that have little or no effect on the response.
4. Second order design phase - this phase starts when the first order design no longer provides a reasonable path of steepest ascent. It augments the first order design with additional points, turning it into a composite design.
5. Ridge analysis phase - if necessary, this guides the search to the optimum using the data from (4).

Note that the user must specify the step size and starting point for each factor.

1.6. Summary

There is a substantial literature on Response Surface Methodology, both in the general statistical area as well as in the computer simulation area specifically. Most papers assume that the metamodel is linear in its parameters, and that the method of least squares is used to estimate those parameters. There are three main approaches to the determination of an experimental design, which I have labelled design property criteria, variance-optimal design criteria, and J-optimal design criteria. Once the experiments have been performed and the model fitted, the model can be analysed using canonical analysis and ridge analysis. To determine the factor settings that optimise the response, there is the choice of path search methods and direct search methods.

Since the process of determining a metamodel is not straightforward, and involves many choices, some research has been reported on developing generic task lists that are intended to support experimenters during the metamodelling process. Van Meel and Aris (1993a,b) note that the literature focuses on statistical rather than procedural aspects. A generic support tool would not only reduce the number of tedious tasks performed by the user, but also allow the user

to focus on procedural rather than statistical aspects. Both van Meel and Aris (1993a) and Tao and Nelson (1994) present such a task list.

There have also been several papers reporting the development of computer software that provides decision support for experimental design in particular. Smith (1976), Gardenier (1990), Hossain and Tobias (1991), and Meidt and Bauer (1992) present such software. However, in each case the software requires the user to choose from a limited number of pre-determined designs (essentially factorial or fractional factorial designs), and as such does not provide expertise.

CHAPTER 2: A CRITIQUE OF CURRENT EXPERIMENTAL DESIGN METHODS FOR SIMULATION

2.1. Introduction

In Chapter 1 it was noted that the variability of the estimates of the metamodel parameters depends directly on the choice of the factor settings $\mathbf{x}_i$ for each experiment, and the proportion p_i (or number n_i) of experiments to be performed at those settings (see equation (1.4)). The set of pairs $\{(\mathbf{x}_1, p_1), (\mathbf{x}_2, p_2), \ldots, (\mathbf{x}_r, p_r)\}$, where r is the number of distinct design points, is referred to as an experimental design. In stochastic simulation, the simulation model has one or more random number streams as inputs. Hence the response of the simulation model is a random variable, and performing a simulation run is indeed experimentation. Experimental design methods may thus be applied to simulation experiments to reduce the variability of estimates of the metamodel parameters. This has been recognised in the simulation literature, and generally standard "classical" designs have been adopted.

Authors of early papers that discussed the application of experimental design methods to simulation studies (for example Hunter and Naylor (1970), Smith (1973b), Biles (1974), Montgomery and Evans (1975), Biles and Swain (1979)), were mostly concerned with encouraging practitioners to use some formal design method, as opposed to ad-hoc or random approaches. The designs most often recommended were the classical factorial, fractional factorial and composite designs. These designs are also used in more recent papers, which have investigated the assumption of constant variance (discussed in section 2.3.) and the addition of random number stream selection into the design process (see for example Schruben and Margolin (1978), Tew and Wilson (1992), Donohue, Houck and Myers (1993a)). One exception is Cheng and Kleijnen (1995), who consider an optimal design approach.

However, the foundations of classical design methods lie in the agricultural context. Mead (1988, p4) comments that

> "If we consider the history of experimental design, then most of the developments have been in the biological disciplines, in particular in agriculture, and also in medicine and psychology. There is therefore an inevitable agricultural bias to any discussion of experimental design."

It is not immediately clear that design methods developed for agriculture are also applicable to stochastic simulation. One important feature of the agricultural context is that experiments are generally conducted *concurrently*. The response data (such as yield) from any experiment is usually collected some time after the experiment is initiated, due to the length of the growing season. On the other hand, simulation experiments are generally conducted *sequentially*. There are a number of further differences between the "classical" and simulation contexts. Yet an extensive search of the literature suggests that there has been virtually no discussion as to the effects of these differences on the application and usefulness of classical design methods in the simulation context.

The focus of this chapter is a critical evaluation of the applicability of classical design methods in the simulation context, and the associated assumptions made in the simulation design literature. In section 2.2. the definition of an experiment in simulation is discussed, while sections 2.3. and 2.4. consider the effect of distributional and sample-size selection assumptions respectively. Section 2.5. considers some further practical restrictions placed on designs. Finally, section 2.6. considers the steps involved in the application of current design methods, and the difficulties of automating such a process.

2.2. The Definition of an "Experiment"

A typical classical experiment involves collecting a *single* observation of a response y_i at design point x_i. For example, in agricultural experiments the observation could be the crop yield for a given experimental plot. Such observations are generally assumed to be independently distributed, although methods have been developed to deal with a small amount of correlation between responses, leading to block designs and randomisation methods (e.g. see Mead

(1988)). Responses are also assumed to have constant variance σ^2 across the design space $\mathcal{X}$. However, although there appears to have been virtually no discussion on this point in the literature, the exact meaning of an "experiment" in simulation may not be a simple analogy to this.

When the simulation model is a terminating model, we have a known stopping condition for each run. For example, this stopping condition could be a given number of customers processed, or the length of time that has been simulated. Hence an experiment for terminating simulations is defined as one run. Examples of responses are the waiting time of the last customer, or the number of customers served during an 8 hour day. Provided independent random number streams are used for each run, the responses are independent. As can be seen, terminating simulations are very similar to agricultural experiments, and it would seem reasonable to apply classical design methods to them.

However, most of the literature on experimental design for simulation concentrates on situations where steady-state conditions are studied. In steady-state simulation we typically collect a large number of observations for each run, so that at a given design point $\mathbf{x}_i$, for run j, we collect the sequence of observations $\{y_{ij1}, y_{ij2}, ..., y_{ijk}, ...\}$. For example, one observation during the simulation of a simple queue would be the time in the system (or some other measure) for a *single* customer. For steady-state simulations the interest generally lies in the mean of those observations, and its variance. Similar to terminating simulations, it would seem logical to define an experiment to be a single run, and to define the response of the experiment to be the mean $\bar{y}_{ij}$ of the observations collected.

A major problem with this definition is that unlike terminating simulations, steady-state simulations do not have an obvious stopping condition. Such a stopping condition, the length of the run, must be chosen (directly or indirectly) by the experimenter. On the other hand, we also cannot define an experiment to be the process of collecting a single observation y_{ij} obtained from the simulation model during a run, and let that observation be the response. This is because generally there is significant correlation between successive observations. This

violates the assumptions made in the classical design literature regarding the independence of the responses.

So in order to allow classical designs to be used in studies of steady-state simulation models, authors of papers in simulation have (implicitly) defined an experiment as a single run. This assumption has been a part of the literature since the first papers on experimental design for simulation.

However as noted above this definition brings a new variable into the problem of experimental design for steady-state simulation: The run length for each run. In turn, this implies that the trade-off between the number of runs performed at each design point and their length should be considered to be part of the problem of finding an appropriate design. This is because depending on the situation, it may be more efficient (in terms of the variance of the mean response) to perform only a few long runs at each design point, or a larger number of shorter runs.

Quantifying this trade-off is usually very difficult however, for two reasons. First, the variance of the response $\sigma^2(\bar{y}_{ij})$ depends on the autocorrelation structure of the response data, and this autocorrelation structure is usually unknown and difficult to estimate. Second, there is the complication introduced by the presence of an initial transient period of variable length for each run, which is usually discarded. So on the one hand, performing only a few long runs means that fewer initial transient periods are required than for a larger number of shorter runs. But on the other hand, in most cases the effect of the autocorrelation structure is such that we would want to keep the run-length for each run as short as possible, and perform many short runs. Whitt (1991) provides an in depth discussion of the trade-off between the number of runs and their length for simulations of queueing models.

Classical design theory clearly does not provide for the determination of the steady-state simulation run-lengths. As a result, it has been (implicitly) assumed in the simulation design literature that these are set, arbitrarily, by the experimenter. This was done to ensure that from the point of view of experimental design, steady-state simulations could be treated as terminating

simulations. Further, in an attempt to satisfy the classical assumption of constant response variance, an additional assumption that is implicitly made in most of the simulation literature is that the run-length is constant across all runs (a few exceptions will be mentioned in section 2.3.).

These three assumptions (that an experiment is one run from which only the mean response is collected, that the run-length is set arbitrarily, and that the run-lengths are equal) allow classical design methods to be applied in the context of steady-state simulation. As a typical example seen in the literature, take a simple 2^2 factorial design, which requires 4 experiments to be performed for each replication of the design. In terms of steady-state simulation, each of the 4 experiments is one run, and we collect only the mean of the observations (the response) from each run. Usually a number of replications of this design are performed, so that more than just 4 responses are obtained.

However, in steady-state simulation the above rigid definition of an experiment effectively forces the experimenter to use what is known as the method of independent replications. This is because regardless of the length of each simulation run, only one response is collected - the mean of the observations $\overline{y}_{ij}$. However there are many well developed methods, such as Batch Means, Spectral Analysis, and Renewal Analysis, which also allow the variance of $\overline{y}_{ij}$ to be estimated from a single run (see Pawlikowski (1990) for an in-depth review and comparison of the advantages and disadvantages of these methods). Although there may be significant efficiency advantages from using these methods, they cannot be applied in conjunction with existing experimental design methods. Note that it may appear that batch means can be used with classical designs, but this is true only if the batch size is known beforehand, which is generally not the case. Without knowing the batch size, we would be unable to relate the number of batches (experiments) to the total run length, and hence not know when to stop the run.

In addition, the use of independent replications constrains the experimenter's ability to choose an efficient mix of runs and run-lengths. The designs seen in the literature most often, such as the above factorial design example, make the possibility of performing only one long run (or just a few) at

each design point very unattractive. In the above factorial design example, if one long run was performed at each design point then only 4 responses would be collected. No information on the estimated variances of these 4 responses is collected. Yet in the case of simulation of the M|M|1 queue, it is most efficient to perform only one long run at each design point (Whitt (1991)), and the use of a method such as Spectral Analysis would provide an estimate of the variance of the response from one run, in addition to the mean.

One paper that does mention variance estimation methods other than independent replications is a paper by Cheng and Kleijnen (1995), who consider an optimal design method for simulation. They recognise that any method can be used to select the mixture of run-length versus number of runs, once the total run-length at a design point has been determined. They also mention the use of variance estimation methods such as spectral analysis to assess lack of fit in cases where only a single run is performed at each design point, but do not discuss the use of such methods in general.

Note that some of these problems are not entirely restricted to simulation experiments. In a classical context a similar problem to the selection of a run-length exists, which appears to have had little attention. Many classical experiments require a decision to be made about the size of a single experiment, such as the size of a plot of land in an agricultural experiment, or the total amount of chemicals mixed together in a laboratory experiment.

2.3. Distributional and Cost Assumptions

One issue that has received some attention recently is the validity of the assumption of constant response variance. This is one of the basic assumptions made in the classical design literature, and is used to justify the application of Ordinary Least Squares (OLS) to find the parameters of the metamodel. Although the parameter estimates will still be unbiased if the response variance is non-constant and OLS is applied, they will not have the smallest variance of all the linear estimators (Draper and Smith (1981). Since the main objective of

experimental design is to ensure that the variances of the estimated metamodel parameters are as small as possible, then it is important that the response variances are indeed constant.

For terminating simulations, the expected run-length at any design point may or may not be constant across $\mathbb{X}$, depending on the stopping condition used. For steady-state simulations the run-length at any design point has been assumed to be constant in the design literature. In both cases, the variance of the response is often a function of the factor settings. As a result, the assumption of identically distributed responses is often not valid in the simulation context, especially when considering simulation models of queuing networks (Whitt (1989)). As an example, Kleijnen, van den Burg and van der Ham (1979) present a case study in which 16 different factor settings lead to mean responses with estimated variances ranging from 64 to 93,228. An example in section 3 of Chapter 5 also illustrates this point.

In general, however, the simulation design literature has continued to make this assumption, often without explicitly mentioning it. Some exceptions are Whitt (1989), Welch (1990), and Kleijnen and van Groenendaal (1995), who report on research into methods that will allow this assumption to be satisfied. The most commonly suggested procedure is to adjust either the run-lengths at each design point (possible for steady-state simulation only), or to consider the mean response at each design point and adjust the number of runs performed to try to achieve constant variance of the mean response. To guide the experimenter in selecting appropriate run-lengths, Whitt presents formulas for the relative run-length required for simple steady-state queuing models, Kleijnen and van Groenendaal present a 2-stage and a sequential procedure for selecting the number of runs, and Welch presents a number of options.

However by modifying either the run-lengths or the number of runs, such procedures modify the experimental design in order to satisfy an assumption. This may have unpredictable effects on the efficiency of the design used. A different approach is used by Cheng and Kleijnen (1995), who consider an optimal design approach and explicitly recognise that the response variance may

be a non-constant function. However, they do assume that the response variance function is known.

The assumption of constant response variance is made not only in the fitting of the model, but also implicitly in the classical designs themselves. For example, a factorial design consists of an equal number of experiments at each design point. Such a design clearly assumes that the response variance is constant, and hence that it is reasonable to sample an equal number of times at each design point. If the response variance was not constant, then it may instead be more efficient to perform more experiments at some design points than others. However, no framework for modifying classical design-property-criteria designs according to the response variances appears to exist.

A related assumption implicitly made by the use of classical designs such as those seen in simulation, is that the cost of an individual experiment (run) is constant across the design region. Again, this assumption is built into most designs, such as a factorial design where an equal number of experiments is performed at each design point. However, in many situations the cost per experiment is not constant. For example, in the Jackson queueing network shown in section 3 of Chapter 5 the factor p influences the proportion of customers that travel through three instead of two nodes (servers). Thus the number of events that need to be scheduled, and hence the CPU time required per run, is very much a function of p.

2.4. Sample-Size Selection

A major assumption made by classical design theory is that the total number of experiments N is given. The experimenter is assumed to know in advance exactly how many experiments will be performed, and as such the design only needs to specify the *proportion* of experiments to be performed at each design point. This assumption is often valid in classical contexts, where the time taken to perform an experiment and the cost of each experiment play an important role. At least one of these is usually considerable, with the result that any experimental

design must remain within a fixed cost or time budget. Thus N is usually known, and relatively small.

However, in the simulation context we almost have the opposite situation. One of the advantages of simulation is that we are able to speed up time, so that the simulation takes far less time to perform than would the collection of the same information from the real system (if this was possible). Also, the computer time required for a simulation is relatively cheap, and is becoming cheaper every day. Hence in simulation the experimenter often does not have a fixed budget for the experiments. Instead, it is likely that the experimenter has some experimental objective based on the amount of statistical information to be collected, such as the accuracy of the metamodel that is fitted to the response data.

In order to allow classical design methods to still be applied to simulation studies, we could set N according to the amount of statistical information collected. However, in most simulation situations the experimenter would be unable to relate the number of runs chosen (and for steady-state simulations their length, which the experimenter is also required to specify) to the amount of statistical information obtained from them. Thus the current literature on design for simulation leaves the experimenter to make a fairly arbitrary sample-size choice. The result is that either too much, or too little information is collected to answer (with suitable confidence) the questions that the simulation study was designed to answer. For example, a possible objective could be the determination of a metamodel with a specified average variance of the fitted response, or specified average confidence interval width. Instead of ensuring that 'sufficient' information is obtained to allow such statistical conclusions to be drawn, the main focus of the simulation design literature has been on design efficiency (maximising information for a given cost) given a fixed sample-size. However it is usually possible to make conclusions on the basis of sufficient data obtained inefficiently, but not on insufficient data obtained efficiently.

When point-estimates are required of the mean response of a simulation model for a given factor setting, there are well documented methods for obtaining, for example, a given confidence interval width (e.g. see Fishman (1971), Kleijnen (1987, Chapter 5), Nakayama (1994)). Such methods assume

that if steady-state results are required, the run-length of each run is known and constant. A simple two-stage procedure takes an initial sample of size n_0, and uses this to determine the total sample-size:

$$N = \left(\frac{t_{n_0-1,1-\alpha/2}}{w} \right)^2 s_y^2,$$

where w is the desired confidence interval half-width, and s_y^2 the estimated variance of the responses. However, these methods do not appear to have been extended to the situation where the experimenter wishes to obtain a *metamodel* with, for example, a given *average* confidence interval width over a design region:

$$\int_{\mathcal{I}} t_{N-1,1-\alpha/2} \sqrt{\mathrm{Var}\left(\hat{y}(\mathbf{x})|N\right)} d\mathbf{x} \Big/ \int_{\mathcal{I}} d\mathbf{x} \leq w.$$

2.5. Further Limitations of Commonly Used Designs

Thus far, a number of undesirable features of standard classical designs such as factorial and central composite design have been highlighted. In this section we show two further features of standard classical designs that restrict their usefulness.

First, in some situations one or more of the factors may be discrete valued, rather than continuous. Often such factors will also have relatively few levels within the region of interest. For example, we may have a situation with two discrete valued factors, one with 4 levels and the other with 5, leading to the grid in Figure 2.1. We now wish to use a central composite design, which allows a second order polynomial model to be fitted. Both the 'star' points (which roughly lie on a diagonal line between the corners of the design region in Figure 2.1.) and the centre point must lie in particular positions for the design to have desired properties such as rotatability, model misspecification robustness, and orthogonality (see Chapter 1). It is clear from Figure 2.1. that this design requires major modification in order to be applied to this situation, and that such

modification would remove many of the properties for which this design was selected.

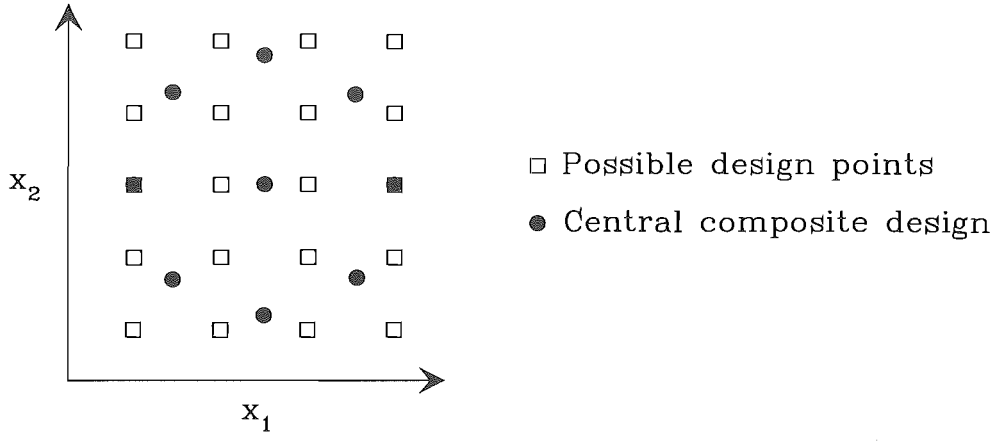


Figure 2.1. A situation with discrete factors

Second, standard classical designs were developed for cuboidal or spherical regions of interest. However it would be quite likely that in some cases, other shapes would be more appropriate (Sargent (1991) lists this as a research issue). For example, in cases where two factors each have a similar effect on the response, then the experimenter may wish to determine the least-cost combination of these factors to achieve a given response. Thus the region of interest could look similar to the area between the lines in the Figure 2.2.

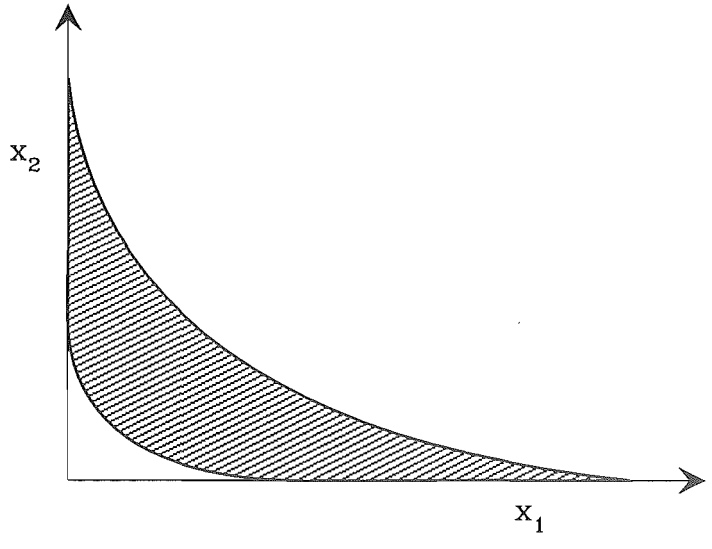


Figure 2.2. A possible design region

Due to the arbitrary nature of factorial-type classical designs, no methods appear to exist which could aid in the modification of standard designs for such situations.

2.6. Difficulties of Automated Design Selection

Computer simulation is the ideal context for which to consider automating the process of designing experiments. For example, on the basis of a small number of inputs, which the experimenter could reasonably be expected to know, computer based algorithms could determine an appropriate experimental design, launch the required simulation runs, and collect and analyse the output. Automated support like this could take the form of a front-end to simulation software. Calls for such software have been made; in particular see Sanchez et al (1994), which contains position statements for a panel discussion at the 1994 Winter Simulation Conference. There, David Kelton stated that (p 1312):

> "While statistical analysis of simulation output data, in post-processing mode, is clearly essential, it is far from the whole story, in my opinion. Practitioners' needs for design-and-control software are just as urgent, and maybe more so. I refer here to a capability that would take a general model (already validated and verified) and a simple description of what to do with it, and will then go do it ...."

In this section we consider the practicality of automating the selection of the components of an experimental design. The components considered to be part of an experimental design (both explicitly and implicitly) in the simulation literature, are:

- $\mathbf{x}_i$ the design points,
- p_i the proportion of designs to be performed at $\mathbf{x}_i$,
- r_{ij} the random number stream used at point $\mathbf{x}_i$ for run j,
- l_{ij} the length of the j^{th} run at point $\mathbf{x}_i$,
- N the total number of runs.

Often the issue of selecting the first three components is referred to as a *strategic issue*, while the latter two components are referred to as *tactical issues* (Kleijnen (1987), Wild and Pignatiello (1991), Sargent (1991), Donohue (1994)).

From the perspective of automation, the automated selection of the basic experimental design (design points and proportions) for arbitrary situations is certainly possible, although probably not satisfactorily. This would first require the selection of an experimental design class such as factorial, fractional factorial or composite simplex, and then the selection of a specific design within those classes. In essence, the problem lies in selecting the design class, as the literature does not provide standard criteria, or rules, for determining which design class is most suitable for a particular situation. Generally the justification given by authors for choosing certain designs for their applications is based on the perceived 'quality' attributed to the design in other literature. Hence any automated approach would either have to use one (and only one) particular design class for any situation, or provide the user with a number of pre-selected options to choose from.

The choice of random number streams for the stochastic components of the simulation model for each run is an issue that affects the efficiency of the design. By appropriately selecting common and/or antithetic random number streams, the estimated variance of the response can be reduced for a given experiment size. There are a number of papers on this topic (see section 2.1.). However, the choice of random number streams is outside the scope of this thesis. A simple although less efficient alternative is to use independent random number streams for each run.

Given the status of current research, the largest difficulty standing in the way of automation appears to be the selection of the run-lengths and total number of runs. Although these choices have been classified as tactical issues, implying that they are of lesser importance than the strategic issues, they are in fact crucial to the success of any simulation study. However there does not appear to be an existing method that would allow the automated selection of these variables,

leaving user-selection as the only alternative. Clearly if the run-length for each experiment and total number of experiments performed are not set appropriately, then one of the two possible outcomes is that not enough information is collected to make a statistically valid conclusion. On the other hand, if the design points, proportions and/or random number streams have not been chosen appropriately, but certain basic rules are followed (e.g. there must be sufficient design points to allow the metamodel parameters to be estimated) and the run-length and total number of runs have been chosen appropriately, then the only effect is that the design will not be as efficient as it could have been.

2.7. Summary

It appears that it is now well accepted that experimental design theory should be used to improve the efficiency of simulation studies. A thorough examination of the simulation literature has revealed that with few modifications, classical experimental design methods have been applied to simulation. Safizadeh (1990, p809) comments that

> "... the statistical designs needed for analysing simulation experiments are quite similar to those used in the physical experiments."

We strongly disagree with such comments or conclusions. In this chapter, we have outlined a number of significant differences between the classical and simulation contexts. The application of classical design methods to simulation has resulted in a number of often inappropriate assumptions being made. This includes the assumptions of constant response variance, constant run-length, and known total number of runs. In addition, such methods are inflexible when it comes to design region shape, discrete-valued factors, and variance estimation methods.

Currently the selection of the components of an experimental design appears to require a number of arbitrary decisions on the part of the experimenter. The literature provides little guidance regarding nearly all the decisions that need to be made in the choice of design. Any front-end software developed for

experimental design on the basis of current methods would thus do little more than perform a number of tedious tasks. Such software would certainly not be capable of making design choices, and provide only marginal decision support. A number of such tools have recently been reported in the literature, e.g. Gardenier (1990), Meidt and Bauer Jr. (1992), Hossain and Tobias (1991).

Together, these observations suggest that an alternative to classical design methods needs to be developed for simulation. Indeed, some authors within the classical literature itself also do not appear to be entirely satisfied with the approach taken there:

> "Most of the important principles of experimental design were developed in the 1920's and 1930's by R.A. Fisher. The practical manifestation of these principles was very much influenced by the calculating capacity then available. Had the computational facilities which we now enjoy been available when the main theory of experimental design was being developed then, I believe, the whole subject of design would have developed very differently. .... Another cause for concern in the development of experimental design is the tendency for increasingly formal mathematical ideas to supplant the statistical ideas. Thus the fact that a particularly elegant piece of mathematics can be used to demonstrate the existence of groups of designs [...] begs the statistical question of whether such designs would ever by practically useful." (Mead (1988, p5))

Figure 2.3. summarises the main points raised in this chapter. The three experimental situations discussed are shown as

- **Agricultural experiments**: The classical agricultural context, where experiments are conducted concurrently.
- **Simulation experiments 1**: Terminating simulations, or steady-state simulations where independent replications is used.
- **Simulation experiments 2**: Steady-state simulation where a variance estimation technique other than independent replication is used.

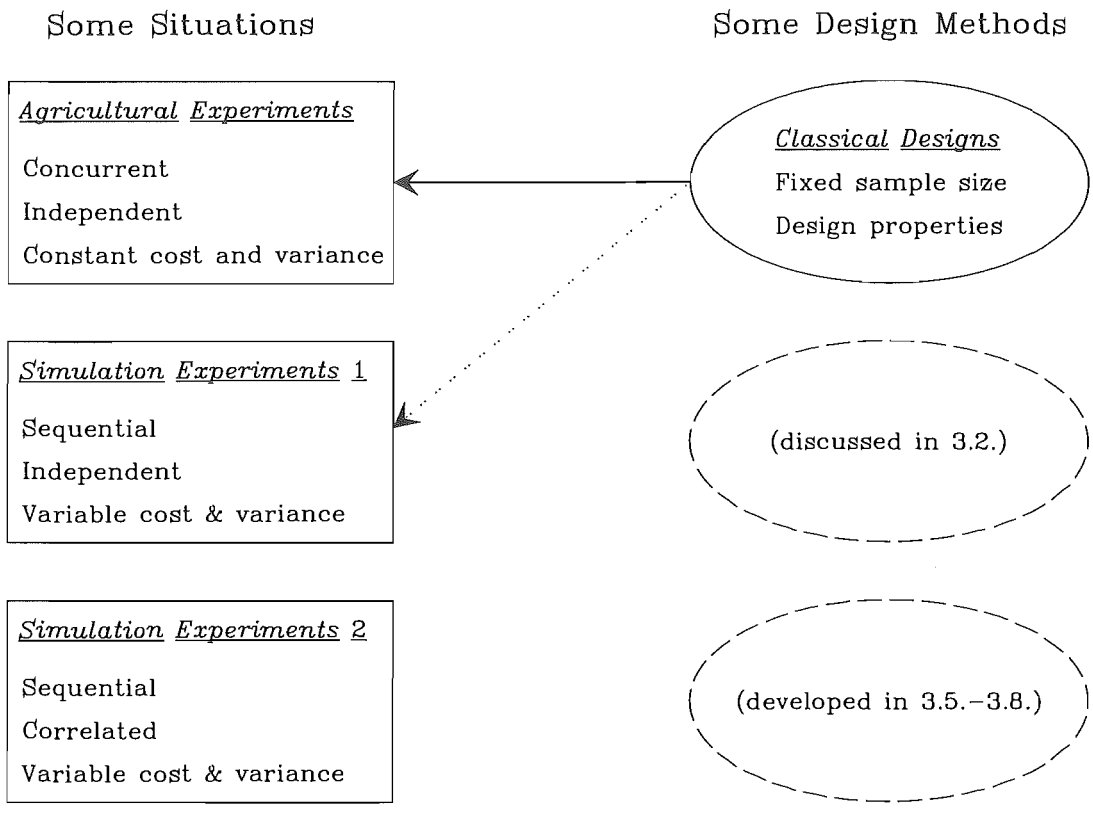


Figure 2.3. Some design methods and experimental situations

Currently there is one basic design approach used in the literature, labelled as classical designs. This approach is suitable for the agricultural situation, but is less suitable for the simulation situation where independent replications is used, and is incompatible with simulation when another variance estimation technique is used. Figure 2.3. will be seen again in later chapters when other design approaches have been developed and discussed.

CHAPTER 3: DEVELOPMENT OF A NEW DESIGN APPROACH

"Semi-sequential, Information Constrained, Optimal Experimental Design"

3.1. Introduction

In Chapter 2 we concluded that classical design methods based on design property criteria are often not appropriate for the simulation context. The application of such methods requires a number of assumptions that are often not valid, resulting in inefficient and inappropriate designs. In addition, such methods are inflexible in a number of respects, and require arbitrary decisions to be made at almost every stage of the design process. These observations suggest that a new design approach should be developed, specifically for the simulation context.

In this chapter, in sections 3.2. and 3.3. we first investigate two existing alternatives to the design methods currently seen in the simulation literature. The first of these is classical optimal experimental design theory. This substantial body of theory emerged in the late 1950's as an alternative to the classical design property criteria approach. The main advantage of optimal design methods is usually stated to be greater design efficiency and a higher level of objectivity in the choice of design. However, until recently optimal design theory has not been part of the simulation experimental design literature (Cheng and Kleijnen (1995)), and its merits in the simulation context do not appear to have been investigated. The second approach investigated is sequential analysis. This is another body of theory that has seen little mention in the simulation design literature, but which would seem to be an obvious approach to overcome the sample-size problem identified in Chapter 2. However, a number of problems are identified for both of these approaches.

In section 3.4 we propose a design approach consisting of a combination of optimal design and sequential analysis. Such a combination overcomes the two main problems faced by these two methods individually, but does not appear to

have been investigated before. However, such a combination retains the other problems, inappropriate assumptions, and inflexibilities of optimal design and sequential analysis.

In the remainder of this chapter, sections 3.5. to 3.9., we develop a new experimental design approach for the simulation context. Our approach overcomes the problems identified for the classical design property criteria, classical optimal design, and sequential analysis approaches. Importantly, our approach allows the development of software that automates the process of experimental design, based on limited input from the experimenter. Because it is possible to closely represent the problem of choosing an experimental design with an optimisation model, the design method for our approach consists of assembling and solving an optimal design problem. It also contains a sequential element.

For this chapter, as in Chapter 2, any reference to an experiment or simulation run implies one run of a terminating simulation model, unless stated otherwise. The extensions to steady-state simulation are discussed in section 9.

3.2. Optimal Experimental Design

Rather than concentrate on attractive design property criteria, as was the focus of traditional research in experimental design, the aim of optimal design theory is to maximise the amount of 'information' obtained from the experiment. This is done by letting the design be the solution to an optimisation problem. Papers by Kiefer (1959) and Kiefer and Wolfowitz (1959, 1960) have provided the main background and motivation for research in this area. Comprehensive treatments of the subject can be found in Fedorov (1972), Silvey (1980), Pazman (1986) and Pukelsheim (1993).

Let the design E be defined as the collection of pairs

$$(x_1, n_1), (x_2, n_2), ..., (x_r, n_r),$$

where r is the number of distinct design points. The first step in an optimal design method is then to choose a function L(E), labelled the design criterion, which is used to evaluate the information content of candidate designs.

The design criterion is usually some function of the Fisher information matrix M of the parameters of the model fitted to the data. For example, the average variance of the fitted response (assuming constant response variance) is given by

$$L(E) = \int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x}) M^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x} \Big/ \int_{\mathcal{X}} d\mathbf{x}$$

$$= \int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x}) \left(\sum_{i=1}^{r} \frac{n_i}{\sigma^2} \mathbf{f}(\mathbf{x}_i) \mathbf{f}^T(\mathbf{x}_i) \right)^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x} \Big/ \int_{\mathcal{X}} d\mathbf{x}.$$

In order to evaluate this criterion, we must have rank(M) = p, where p is the number of parameters in the metamodel. This implies that the number of distinct design points r must be greater than or equal to p. For some design criteria, designs with $r < p$ may be provide better design criterion values. However, we will assume that such designs are not acceptable, since they do not allow all of the parameters of the metamodel to be estimated (this will be discussed further later). Note that we wish to minimise the value of most design criteria, since most are some measure of variability.

The next step is to assemble the design problem:

$$\begin{aligned} \text{Min} \quad & L(E) \\ \text{s.t.} \quad & \sum_{i=1}^{r} n_i = N \\ & \mathbf{x}_i \in \mathcal{X} \quad \forall i \\ & n_i \geq 0 \quad \text{and integer} \quad \forall i \end{aligned}$$

which can be solved to obtain the optimal design. Because this is an optimisation problem, the solution to this problem is the most efficient design (the design that minimises L(E) given N) for the criterion used. This is in contrast to design property designs, which simply demonstrate a number of desirable combinatorial-type properties. Kiefer and Wolfowitz (1959) illustrate the difference in efficiency through an example of fitting a cubic model over [-1, 1], where the

design criterion is the variance of the coefficient of x^3. In that case, a four point factorial design requires 38% more observations than an optimal design to obtain the same criterion value.

One difficulty with solving the design problem is the integer restriction on the n_i's, which together with a nonlinear objective function makes it difficult to find a solution. Kiefer and Wolfowitz (1959) removed this restriction by redefining the experimental design in terms of the proportion p_i of experiments performed, rather than the number n_i. The resulting designs are known as continuous designs, and need to be rounded to give integer designs.

In terms of the issues raised in Chapter 2, optimal designs have a number of advantages over designs based on design property criteria. Most importantly, the use of a design problem implies a modelling approach, where the experimental situation is modelled to obtain a 'best' solution to the problem of finding a design. Hence an optimal design is chosen using a more objective approach than is used in the rather arbitrary approach of choosing between, for example, factorial and composite designs. As a result, it is possible to automate the process of determining the basic design using optimal design methods, once a number of inputs have been specified by the experimenter. Also, such a modelling approach allows specific experimental situations to be taken into account, such as various design region shapes. Any convex design region can be used by simply expressing it in terms of a set of constraints on the $\mathbf{x}_i$ in the design problem.

However optimal design theory is still based on the same classical assumptions as design property criteria methods. In particular, most of the optimal design literature makes the assumption that the variance of the response and the cost per experiment are constant. Atkinson and Cook (1993, p2) note that

> "All of the substantial literature on optimum designs for linear and nonlinear models assumes additive errors, usually of constant variance. .... In the design literature, the possibility of additive but heteroscedastic errors, in particular, has been considered mostly in the case where the variances are known up to a proportionality constant."

As the quote suggests, a few authors, mainly of early papers and comprehensive texts, do discuss the possibility of heteroscedasticity and variable cost. The suggested procedure of dealing with this is to transform both the response and the factors, so that if $c(\mathbf{x})$ is the cost-per-experiment function and $\upsilon^2(\mathbf{x})$ the variance function of the response, then a scaling factor of $c(\mathbf{x})(\upsilon^2(\mathbf{x}))^{1/2}$ is used (e.g. see Kiefer and Wolfowitz (1959), Cheng and Kleijnen (1995)). The proportions p_i are then interpreted as the proportion of the total budget spent at point i. However this assumes that both $c(\mathbf{x})$ and $\upsilon^2(\mathbf{x})$ are known up to a constant of proportionality, and that the experimenter has a fixed (and known) budget for the experiments. Atkinson and Cook (1993) extend optimum design theory to allow for a parametric structure in the variances, but the resulting designs depend on unknown parameters, thus requiring a Bayesian approach.

Although the determination of an optimal design may seem to be an objective process, it requires selection of a design criterion. The types of design criteria shown in the literature, such as D-, G- and A-optimality, do not appear to be aimed at any particular situations. Rather they are deemed to be useful because of their general statistical properties. Again, this implies that for any particular situation the arbitrary selection of a criterion is required, providing a stumbling block for automation. This is very similar to the problem of choosing between the factorial and composite simplex design classes.

Finally, the assumption that N is known is still made in the optimal design literature. One procedure that is sometimes briefly mentioned is to choose an N that is associated with an acceptable value for the design criterion. However there are two problems with this suggestion. First, the values of the design criteria seen most commonly in the design literature are not easily interpreted in this way. For example, the value of D-optimality is related to the volume of a confidence region for the metamodel parameters. Second, without a sequential element this procedure assumes that the variance function of the response and the cost-per-experiment function are known, or at least that very good estimates are available.

Cheng and Kleijnen (1995) appears to be the only paper that investigates optimal design theory in a simulation context. The main difference between their approach and the approach taken in the rest of the simulation literature, besides

the optimal selection of the design, is that they recognise that the experimenter is free to choose the mixture of runs and run-lengths for steady-state simulations, once the total run-length has been fixed. However they assume that the variance function of the response, for the run-length chosen, is known. They also define an experiment as a single simulation run, and collect only the mean response from each run. Hence their approach cannot directly be applied to steady-state simulations where a variance estimation method other than independent replications is used.

Optimal experimental design can be applied to situations where experiments are conducted concurrently or sequentially, and is able to take non-constant variance and cost-per-experiment functions into account. However, it appears that classical optimal designs are only suitable for terminating simulations, or steady state simulations where the method of independent replications is used. This is shown in Figure 3.1., which updates Figure 2.3.

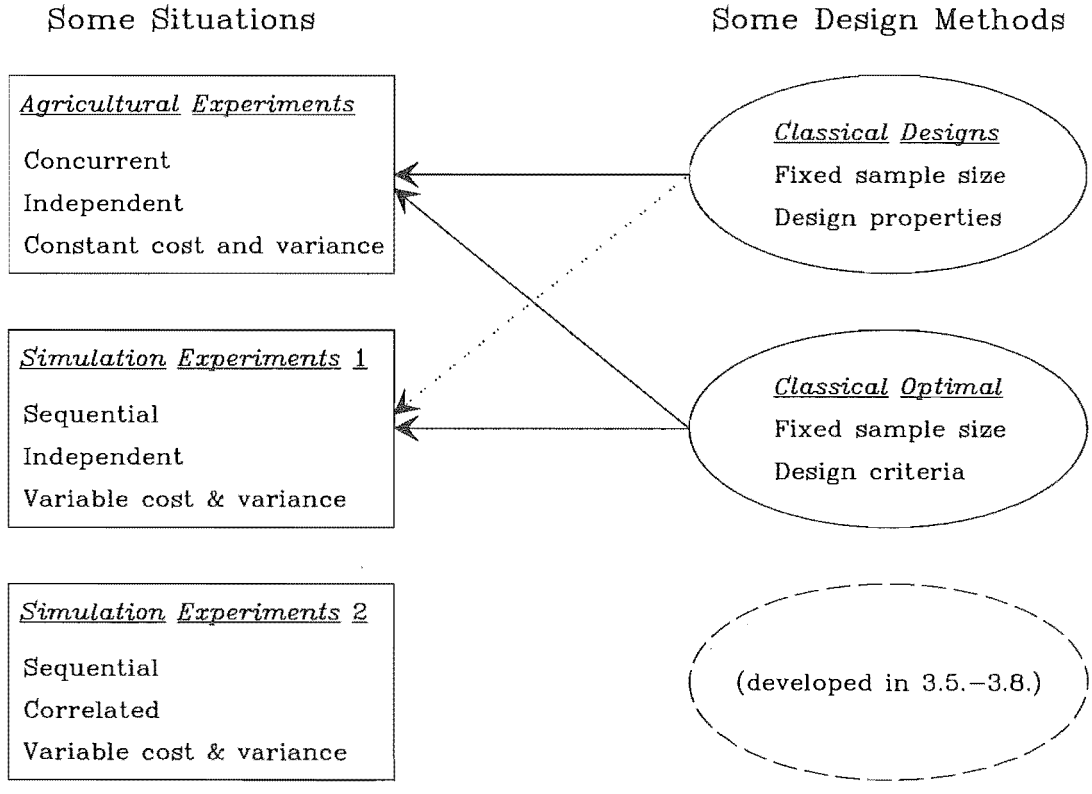


Figure 3.1. Some experimental situations and design methods

3.3. Sequential Analysis

The second body of literature that contains methods that could overcome some of the problems with classical design methods, is that concerning sequential methods. These methods specifically attempt to tackle the problem of finding a suitable sample size. A number of researchers have stated that future research in simulation metamodelling will need to incorporate elements of sequentiality. For example, Welch (1990, p 394) concludes that

> "... in the future we will see closer and closer coupling of statistical calculations and simulation run control. Sequential procedure will be applied not only to model fitting but also to generating model coefficient confidence intervals of fixed widths, to model and parameter selection, to adaptive response surface estimation, etc. Only by the close coupling of statistics and run control can efficient and effective application of simulation be made. There is a deep need to investigate these issues and to build appropriate high level control into simulation packages."

Ghosh and Sen (1991) and Chernoff (1972) provide reviews of most of the methods in the sequential analysis literature. However, the regression context does not appear to have been a significant part of the research on sequential methods, which tends to focus on sampling at a single point or the comparison of two points. A relatively small number of papers (Gleser (1965), Albert (1966) Srivastava (1967,1971)) and Chaturvedi (1987) present sequential procedures designed to produce a fixed-width confidence interval for regression parameters with prescribed coverage probability. These procedures require the existence of a sequence of vectors $\{\mathbf{x}_1, \mathbf{x}_2, ...\}$ representing the factor settings for each successive experiment. Sequentially, experiments are then performed at the design points in this sequence, until a given stopping condition has been reached.

Although these sequential procedures eliminate the problem of needing to select the sample size before experimentation starts, they have a number of features that may not be desirable in the simulation context. First, such procedures require the experimenter to determine a sequence specifying the order

in which the experiments are to be performed. For example, for a design measure taking the value $1/3$ at each of the design points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, one possible sequence is $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots\}$. A problem that immediately springs to mind is how one would go about turning a design measure like (0.345, 0.523, 0.868) into such a sequence, and how well the original design is preserved in this process.

Second, a practical problem with performing experiments using such a procedure is that the experimenter will frequently need to change the factor settings, possibly as often as after each experiment (as in the example above). For example, in a laboratory situation this may require adjustments to equipment used in the experiments. In a simulation context, changing the factor settings for each run is relatively simple. However, the properties of sequential methods such as those found in Gleser (1965), Albert (1966) and Srivastava (1967,1971) are generally asymptotic properties and thus the application of those methods requires that a substantial number of experiments are performed. In simulation, this requires that a large number of runs be performed. In turn, for steady-state simulation this implies that the experimenter must choose a sufficiently small run-length to ensure that the total run-length remains reasonable. Once again, this restricts the experimenter's ability to choose an efficient mix of runs versus run-length.

Finally, sequential analysis is based on a framework similar to classical design methods, so that the definition of an experiment is the same. It thus inherits most of the problems already identified for classical designs. In particular, the choice of design is arbitrary, and since only one number is collected from each experiment the method of independent replications must be used in steady-state simulation.

There are many alternative sequential procedures that could be used, such as procedures where each stage consists of one or more complete repetitions of the design, rather than only a single experiment as assumed above. One such an approach is given in Donohue, Houck and Myers (1993b), which appears to be the only paper on sequential estimation of a metamodel in simulation. They

assume that the metamodel is either first or second order, and propose a sequential approach with a pilot and three stages as follows:

- *Pilot experiment*: Consists of a 2-level factorial design with replicated centrepoints, to allow checking of assumptions such as constant variance and the correlation structure (used in the first and second stage),
- *First stage*: Consists of a fractional factorial design with replicated centre points, to provide information that allows optimal determination of the design points for stage 2,
- *Second stage*: Consists of a factorial design with the position of the points determined optimally using the J-optimal approach (see Chapter 1), to estimate a first-order metamodel and test for curvature,
- *Optional third stage*: Axial portion of a central composite design (augments the stage 2 factorial design), to allow estimation of a second-order metamodel if curvature was detected at stage 2.

However, this approach is designed to provide at each stage information for the optimal determination of the design for the next stage, and is not concerned with sequential sample-size determination.

3.4. Combining Optimal Design and Sequential Analysis

The optimal design and sequential analysis literature for the problem of estimating a metamodel appear thus far to have been segregated. The main assumption of the former is that the overall sample-size is known, while the latter assumes that the experimental design is known. One approach to the complete process of experimentation would be to use optimal design theory to determine the design, and sequential analysis procedures to determine the required number of replications of this design.

However such a combination removes only two of the many problems of the individual approaches, namely design selection and sample-size selection. Also, such a combination may produce unexpected results (as, in fact, optimal design will also do). This can be seen as follows. As discussed in Chapter 2, in many situations the variance of the response is not constant over the design region. Usually the experimenter has some, but not complete, knowledge of this variance function. Hence the 'optimal' design is only optimal with respect to the variance function used to determine that design. Now the experimenter may anticipate before experimentation takes place that, besides the sequentially attained G-optimal value (for example), the design used will also have particular secondary characteristics, which are not formally included in the design problem. However this anticipation may prove incorrect if the true variance function is significantly different from that expected.

For example, take a common two-point G-optimal design for fitting a first order polynomial in one factor, based on (assumed) constant response variance. This design requires that an equal number of experiments be performed at each design point, where the design points lie at opposite ends of the design region. We would anticipate that the value of the variance of the fitted response over the design region would be symmetrical around the mid-point of this region. However if the variance of the response happens to be larger at one end of the design region than the other, then application of the design will result in a fitted response variance function that is skewed. An example in section 1 of Chapter 5 illustrates this effect.

Table 3.1. summarises some important features of the approaches discussed thus far.

[♦ means a problem exists (♦) means the assumption is generally made] *Issue*	*Design property criteria*	*Optimal design*	*Sequential analysis*	*Optimal design & sequential analysis*
Automation issues				
Arbitrary choice of design	♦		♦	
Sample size (budget) assumed given	♦	♦		
Assumptions made				
Constant response variance	♦	(♦)	n/a	(♦)
Constant cost per experiment	♦	(♦)	n/a	(♦)
Inflexibility issues				
Runs vs run-length mix	♦	♦	♦	♦
Variance estimation methods	♦	♦	♦	♦
Design region shapes	♦		n/a	
Practical issues				
Unexpected information distribution	♦	♦	n/a	♦
Design point sequence problem			♦	♦

Table 3.1. Comparison of three approaches

3.5. Sketching Out a New Approach

Table 3.1. shows that current design methods, as well as the proposed combination of optimal design and sequential analysis, have a number of problems when applied to simulation. In the remainder of this chapter we develop a new experimental design approach for the simulation context, which overcomes these problems. This approach is based on the classical optimal experimental design approach, and incorporates an element of sequentiality.

To begin with, we assume that the response, factors and metamodel are as defined in Chapter 1. We also initially assume that the response has constant variance across the design region, that the cost per experiment is constant, and

that the simulation model is a terminating model. These assumptions will be relaxed at various stages of the development of our approach.

By considering classical design approaches, it is clear that the research behind them has in mind a particular type of application. First, the total number of experiments N is assumed to be fixed, known, and relatively small. The latter is shown by the emphasis, most clearly shown in alphabetic optimality methods, on obtaining exact integer replication designs (e.g. Cook and Nachtsheim (1980)). This suggests a type of experiment that is very costly to perform in terms of either time or other resource costs, and hence N is small. Indeed, the authors who established RSM as a methodology, Box and Wilson (1951), would have faced exactly that type of experimental scenario. Both worked at Imperial Chemical Industries, and state typical responses as being yield, purity and cost, and typical factors as temperature, pressure, time of reaction, and proportions of the reactants. Similarly, Fisher's original application of experimental design (see Fisher (1990)) was to agricultural experiments. Experiments with such factors and responses, where a sizeable physical experiment is performed that may take considerable time, would certainly be limited to have a relatively small N.

Second, it is generally assumed that the error variance of each experiment σ^2 is constant. With a small N, it may often be difficult even at the end of all the experiments to estimate σ^2 to a reasonable accuracy.

By considering the classical approach to optimal experimental design, we see that the design problem for the classical context consists of finding a design made up of the variables

$\{\mathbf{x}\}$, the factor settings
$\{p\}$, the proportion of experiments conducted at each $\mathbf{x}$,

by solving the problem of finding the design that minimises a design criterion.

However, the context of terminating simulations is quite different from the above scenario. For such simulations, the experimenter expects to perform a

relatively large number N of experiments. In contrast to many classical experiments, each simulation experiment generally takes relatively little time and costs little to perform. Thus for terminating simulation models it is common to perform in excess of 5 to 10 runs at each design point. As there are many repetitions at each design point, a good estimate of σ^2 can be found from the data.

But most importantly, the emphasis in this type of experimental application is often on making sure that "sufficient" data is collected to make the experimentation worthwhile, rather than staying within a fixed budget. Hence the experimenter does in general not know beforehand what the value of N is.

For such an application, the experimental design shown above and its associated design problem as described earlier would require the experimenter to arbitrarily select the size of N, and then allocate this to the various experimental points according to the proportions specified by the optimal design. But in the context of terminating simulations, it makes little sense to go to some effort to find an optimal experimental design, and subsequently to potentially eliminate any gain made over an arbitrary design by not paying similar attention to the size of the experiment N. Efficiency considerations are important, but the choice of design influences only *how much information we obtain from a set N*. On the other hand, *it is N that determines the maximum amount of information obtainable*. So for terminating simulations we need the experimental design,

{x}, the factor settings
{n}, the *number* of experiments performed at each x,

and a suitable design problem to determine the optimal design.

A general form for a suitable design problem can be found as follows. Assume that the experimental points have been determined. The question now is: How many experiments to conduct at each experimental point? In general, for simulation the answer should be as many as are necessary to allow the required conclusion to be drawn. This required conclusion, or knowledge goal, can be defined as some condition to which the data collected must conform, and labelled

as a target value for the design criterion. So on the one hand, we want to conduct at least as many experiments as are necessary to achieve the knowledge goal. On the other hand, we obviously do not want to conduct any more experiments than are necessary, by being both efficient and not exceeding the goal. Returning to the problem of finding an appropriate design for simulation experiments, this suggests the following new approach to optimal experimental design: *The optimal design should minimise the experimental cost (effort), while satisfying a specified knowledge goal.* In contrast, the classical optimal approach minimises the value of a design criterion, while satisfying a specified cost goal.

For terminating simulations, the combination of using the above definition of the design and design problem will result in an optimal design that provides the experimenter with all the information needed to perform the experiments. A derivation of the mathematical form of the above design problem can be done through consideration of the experimental loss function, which is discussed next.

3.6. The Loss Function

Although there is a substantial literature on optimal experimental design theory, little appears to have been written on the justification for the form of the classical optimal design problem. Indeed, the paper by Kiefer and Wolfowitz (1959) that made optimal design theory popular does not present such a justification, and neither do more recent texts on the subject (e.g. Silvey (1980), Pazman (1986), Pukelsheim (1993)). Almost all of the literature simply assumes that the design problem consists of the minimisation of some design criterion, given a fixed number of experiments. As a result, some of the fundamental assumptions of this design method have received limited attention.

One exception is Fedorov (1972), who uses the concept of an experimental loss function to justify the form of the design problem. This section will expand on the work of Fedorov, and provide the motivation for an alternative design problem.

We can view an experiment as a 'black box', which requires inputs and produces an output, as in Figure 3.2.

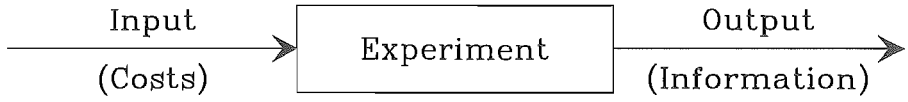


Figure 3.2. An experiment as a black box

The inputs can include physical resources such as fertilisers in an agricultural experiment or computer time in a simulation experiment, as well as other inputs such as labour costs. Some of these inputs may vary with the number of experiments, while others are fixed. Inputs required for an experiment can be labelled as experimental 'costs'. On the other side of the black box, output is obtained, consisting of measurements of the response. Usually these responses are combined in some way to obtain some measure of 'information'.

Clearly the experimenter would like to minimise the experimental costs, while maximising the information obtained. The resulting trade-off can be shown as in Figure 3.3.

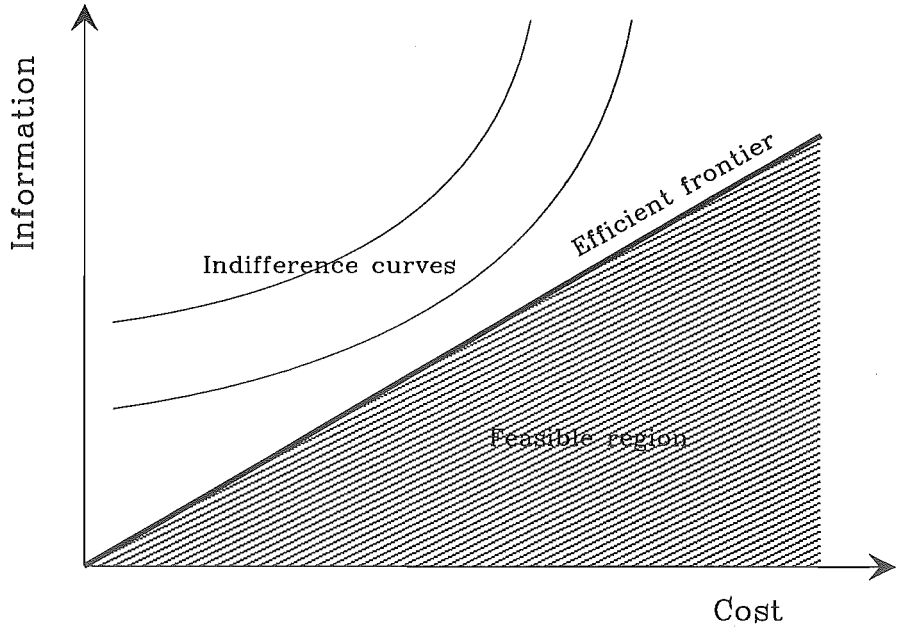


Figure 3.3. The trade-off between cost and information

In the cost-information space the *feasible region* shows the combinations of cost and information that are possible. Because in general an experiment can be performed at less than maximum efficiency, a range of information is associated with any level of cost. The points associated with the highest possible level of information for a given cost lie on the *efficient frontier*. Clearly the experimenter wishes to choose a cost-information point that lies somewhere on this frontier. The exact choice depends on the experimenter's *indifference curves*, which indicate the relative importance placed on the two opposing considerations of maximising information gained while minimising the costs incurred. Note that when using the measures of information used most often in the optimal design literature, and a cost function linear in the number of experiments performed, the efficient frontier is indeed a straight line (this will be shown later). However, in general the efficient frontier may take a number of shapes, and may even be discontinuous.

In general, the experimenter's preferences are difficult to quantify, especially since cost and information are measured in different units. The aim of experimental design theory then is to help the experimenter choose a combination of information and cost that (at least) lies on the efficient frontier. This can be done by modelling the trade-off between the two opposing objectives using an experimental loss function.

First, define an experimental design E as the collection of controllable variables that determine the cost incurred and information obtained from the experiment. Generally the information obtained from an experiment is considered to be inversely proportional to some measure of the variability of the responses. Let $\Psi(E)$ be the 'loss' resulting from the variability of the responses, and $\tau(E)$ be the 'loss' resulting from the cost of the experiment. Then an additive linear experimental loss function is

$$R(E) = \tau(E) + \Psi(E).$$

The optimal design E^* is then the design that minimises $R(E)$.

More specifically, define an experimental design E to mean the collection of pairs

$$(x_1, n_1), (x_2, n_2), \ldots, (x_r, n_r)$$

where r is the number of distinct design points x_i at which the number of experiments performed, n_i, is greater than zero. At this stage we assume that the cost of any experiment is a constant c for $x_i \in \mathcal{X}$. Hence the first part of the loss function is given by

$$\tau = cN(E) = c\sum_{i=1}^{r} n_i .$$

For the second part of the loss function, define L(E) to be a *design criterion* that measures variability of the response or some function of the estimated model parameters, and k a normalising constant. Thus

$$\psi = kL(E).$$

In the optimal design literature, the design criterion is generally based on the Fisher information matrix M of the model parameters. Note that if L(E) is a linear function of M, then $c\Sigma n_i \propto 1/L(E)$, leading to a straight line efficient frontier as shown in Figure 3.3. One example of L(E) is the average variance of the fitted response, given by

$$L(E) = \int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x}) \left(\sum_{i=1}^{r} \frac{n_i}{\sigma^2} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1} \mathbf{f}(\mathbf{x}) d\mathbf{x} \Big/ \int_{\mathcal{X}} d\mathbf{x}.$$

This criterion is suitable when the model is to be used for prediction purposes, when the experimenter does not know in advance which part of the design region the predictions will be made for. Note that the matrix inverse of M within this criterion requires that rank(M) = p. For some criteria, such as criteria based on the gradient of the metamodel at a point (which does not depend on the intercept parameter), the optimal design may have fewer distinct design points than there are metamodel parameters. However, we will assume that in all cases the

experimenter also wishes to estimate all the parameters of the metamodel, so that $r \geq p$.

The loss function can then be written as

$$R(E) = cN(E) + kL(E).$$

The experimental design problem thus consists of finding the design E^* that minimises $R(E)$, being a weighted sum of the number of experiments performed and the design criterion value associated with the design. However to obtain E^* we are required to provide a value for the ratio c/k. This involves consideration of the trade-off between the number of experiments we wish to perform, and the amount of statistical information we wish to collect. In general this is a very complex - and often non-linear - trade-off, and the experimenter is unlikely to be able to provide an appropriate value for c/k.

3.7. Derivation of the Design Problem for the New Approach

In this section, the experimental loss function is used to derive both the classical optimal design problem, and the first stage of the design problem for the new approach. In the previous section it was noted that it is in general not possible to minimise the loss function directly, because the ratio c/k is not known. However, by simply subtracting a constant cN_0 from the loss function, we obtain the modified loss function

$$R'(E) = c\left[N(E) - N_0\right] + kL(E).$$

This loss function is the Lagrangian of the following optimisation problem:

$$\begin{aligned} \text{Min} \quad & L(E) \\ \text{s.t.} \quad & N(E) = N_0 \\ & \mathbf{x}_i \in \mathcal{X} \quad \forall i \\ & n_i \geq 0 \quad \text{and integer} \quad \forall i \end{aligned} \tag{3.1}$$

which is precisely the basic classical optimal design problem. Thus by making the assumption that the number of experiments is fixed, i.e. $N(E) = N_0$, the problem of finding E^* is transformed into a non-linear optimisation problem, or *design problem.*

Note that the design problem shown above is not in the form in which it is most commonly seen in the literature. Usually it is simply stated as "Minimise L(E)", with the implicit assumption that the number of experiments is fixed. As a result, the role that the cost function plays in the design problem has largely not been considered. In fact, a large proportion of the literature deals with the design E^c, consisting of the collection of pairs

$$(\mathbf{x}_1, p_1), (\mathbf{x}_2, p_2), \ldots, (\mathbf{x}_r, p_r)$$

where $p_i = n_i/N_0$. The result is that the design problem only has the constraint $\Sigma p_i = 1$, and the original assumption that the cost of each experiment is constant is no longer 'visible' in the design problem.

The classical approach is suitable for those situations where N_0 is known in advance. However in many experimental situations such as simulation, a more realistic and important practical consideration is to focus on obtaining results to within a certain accuracy, as measured by the design criterion. We can then consider that a target or acceptable level for L(E) has been set, say L_0. The loss function then becomes

$$R''(E) = cN(E) + k[L(E) - L_0],$$

which is the Lagrangian of the following design problem:

$$\begin{array}{lll} \text{Min} & cN(E) & \\ \text{s.t.} & L(E) \le L_0 & \\ & \mathbf{x}_i \in \mathcal{X} \quad \forall i & \\ & n_i \ge 0 \quad \text{and integer} & \forall i \end{array} \tag{3.2}$$

This is the basic mathematical form of the design problem for our approach described in section 5 of this chapter.

The design problem (3.2) will clearly produce the same optimal design as the classical design problem (3.1) if L_0 and N_0 respectively are set appropriately, because they are derived from the same Lagrangian. However there are several important differences between them. First, by minimising the number of experiments performed, but still requiring that a target value L_0 for the design criterion is met, a design that is optimal for (3.2) is both *efficient*, i.e., $L(E^*)$ is the maximum $L(E)$ attainable given that $N(E^*) = N_0$, and *economical*, i.e., $N(E^*)$ is the minimum $N(E)$ required to achieve $L(E^*) = L_0$. Rather than arbitrarily specifying the number of experiments N_0, the design problem allows $N(E^*)$ to be determined optimally. Second, implicit in (3.2) is the requirement that the design criterion used is such that the experimenter can set an appropriate target for it. An optimally designed experiment should take account of the specific objective of the experiment, through the choice of design criterion. This contrasts with the design criteria seen most commonly in the literature, which are usually based on statistical properties which are deemed desirable in a general context.

There are many authors who have also stated that the objective of optimal experimental design is to allow the experimenter to reach a conclusion with minimum cost, but inevitably this is followed by the statement "let N be given". However it appears that there are two proposed methods that would seem to resemble the new approach to experimental design.

In the first, Fedorov (1972, p61) notes that

> "In many experimental investigations, an 'accuracy' of determination of the estimates of the sought parameters is given beforehand.",

and in such a case he advocates adjusting N until a design $E(N+1)$ is found such that

$$\min L(E(N)) > L_0 \geq \min L(E(N+1)),$$

where $E(N)$ denotes a design consisting of N experiments. This is an iterative approach of setting N and determining $\min L(E(N))$. However, this approach

considers only part of the loss function R(E) and so does not consider the costs of experimentation.

Second, Mitchell (1974) briefly mentions a sequential option in his computer program that allows the user to approximately set N in order to "achieve satisfactory precision" (p206). The resulting design is not optimal, and needs to be optimised by his DETMAX procedure which then D-optimises the design for the specified N.

3.8. Adding a Sequential Element

The new design problem (3.2) as developed thus far overcomes one of the main limitations of the classical approach - that the total number of experiments to be performed needs to be known in advance. Two crucial assumptions have been made to achieve this: (i) That the response variance is constant across $\mathcal{X}$ and known exactly, and (ii) that the cost-per-experiment is constant. If this is the case, then the design problem (3.2) correctly models the experimental situation. The cost function value of the optimal design E*, and its design criterion value L(E*), will then be equal to the expected cost and design criterion value of the response data collected.

However, in most experimental situations the cost-per-experiment and response variance functions are neither constant nor known exactly. In addition, the limitations resulting from the definition of an experiment in simulation are still present in the design problem (3.2). An experimental design is still defined in terms of the number of experiments (runs) performed, and hence our approach as developed thus far is not suitable for steady-state simulations where a variance estimation method other than the method of independent replications is used.

In this section the experimental design is redefined, leading to a design problem that does not require these assumptions, and a design that is suitable both for terminating simulations and for steady-state simulation in combination with any variance estimation method. This is achieved by transforming (3.2) into a design problem that is non-sequential, but which has as solution an

experimental design consisting of stopping rules similar to those seen in sequential analysis.

We now consider terminating and steady-state simulation separately.

Terminating Simulations

First we consider those situations where the simulation model is a terminating model. Hence only one response is collected per experiment (run), and responses are independent due to the (assumed) use of different random number streams for each run.

The essence of the transformation lies in re-defining the experimental design. Instead of the number of experiments n_i performed at a design point $\mathbf{x}_i$, we focus on the variance of the mean response, $\mathrm{Var}(\bar{y}(\mathbf{x}_i)|n_i)$, to be obtained by those experiments in some sequential fashion. By an experimental design E^{σ} we will mean the collection of pairs

$$(\mathbf{x}_1, \sigma_1^2), (\mathbf{x}_2, \sigma_2^2), \ldots, (\mathbf{x}_r, \sigma_r^2)$$

where we define

$$\sigma_i^2 = \mathrm{Var}\left(\bar{y}(\mathbf{x}_i)|n_i\right) = \frac{\mathrm{Var}\left(y(\mathbf{x}_i)\right)}{n_i} = \frac{\upsilon^2(\mathbf{x}_i)}{n_i}. \tag{3.3}$$

Note that we have now relaxed the assumption of constant variance. Since n_i must be integer, then σ_i^2 can also only take on certain values, given by the right hand side of (3.3).

We can now assemble the design problem that we will solve for the optimal design $E^{\sigma *}$. Most design criteria are based on the Fisher information matrix, M, of the model parameters. The classical form of M,

$$M(E) = \sum_{i=1}^{r} \frac{n_i}{\sigma^2} f(\mathbf{x}_i) f^T(\mathbf{x}_i),$$

where σ^2 is the (assumed) constant variance of the response data, is easily changed to the Weighted Least Squares estimator

$$M(E^{\sigma}) = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} f(x_i) f^T(x_i),$$

using (3.3). This allows most design criteria to be expressed in terms of the new design E^{σ}. Similarly, the experimental cost function

$$\tau(E) = c\sum_{i=1}^{r} n_i,$$

becomes

$$\tau(E^{\sigma}) = \sum_{i=1}^{r} \frac{c(x_i)\upsilon^2(x_i)}{\sigma_i^2}.$$

(note that the cost per experiment is now a function of x_i). We then obtain by a similar argument used to obtain (3.2) the modified design problem

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{r} \frac{c(x_i)\upsilon^2(x_i)}{\sigma_i^2} \\ \text{s.t.} \quad & L(E^{\sigma}) \le L_0 \\ & x_i \in \mathcal{X} \quad \forall i \\ & \sigma_i^2 = \upsilon^2(x_i)/n_i \quad \forall i \\ & n_i \ge 0 \quad \text{and integer} \quad \forall i \end{aligned} \tag{3.4}$$

where $c(x_i)\upsilon^2(x_i)$ can be seen as the cost of obtaining a unit of $1/\sigma_i^2$.

The design problem (3.4) is essentially no different from the design problem (3.2). However, we can obtain a significantly more useful design problem simply by relaxing the integer requirement on n_i. The result of this is that the σ_i^2 are no longer restricted to only take on certain values. This leads to the design problem

$$\boxed{\begin{aligned} \text{Min} \quad & \sum_{i=1}^{r} \frac{c(x_i)\upsilon^2(x_i)}{\sigma_i^2} \\ \text{s.t.} \quad & L(E^{\sigma}) \le L_0 \\ & x_i \in \mathcal{X} \quad \forall i \\ & \sigma_i^2 \ge 0 \quad \forall i \end{aligned}} \tag{3.5}$$

This design problem represents a 'hybrid' approach to experimental design, that lies between the classical optimal design and sequential analysis approaches. The objective of this approach can be described as the determination, non-sequentially, of a set of design points with associated stopping rules, which will minimise the experimental cost required to ensure that a target value for the design criterion is met. Although the design is determined non-sequentially, experimentation at *each design point* is continued until a stopping rule like

$$n_i = \min\left\{n: n \geq n_0, s_i^2(n) \leq \sigma_i^2\right\}, \tag{3.6}$$

where $s_i^2(n)$ is the estimated variance of the mean response after n experiments, has been satisfied. Thus we label this approach as "Semi-sequential, Information Constrained, Optimal Experimental Design", or SICOED.

The variable σ_i^2 of the SICOED design problem is not restricted to take on only certain values corresponding to an integer number of experiments. The optimal values of σ_i^2 will then (in theory) correspond to a non-integer value of n_i, which is effectively rounded up by the stopping rule (3.6). Thus the optimal design for the SICOED design problem (3.5) will not be the optimal design for (3.4). In fact, the optimal design for (3.5) will be slightly less efficient and economical for the design problem (3.4) than the optimal design for (3.4) will be. The integer nature of experiments means that the number of responses collected is equal to, or slightly greater than, required by the design. Thus the design criterion value calculated from the actual responses will be slightly better than required by its target. When the number of experiments is sufficiently large, these effects become insignificant.

Note that use of the SICOED approach is restricted to those experimental situations where experiments are conducted sequentially, rather than concurrently. This is because the focus lies on the variance of the mean response at each design point, rather than the number of observations. As a result, the SICOED approach can be used in many laboratory and simulation situations, but not in the classical agricultural context.

Also, if we set $c(x_i)v^2(x_i)$ to be a constant, then - by setting L_0 and N_0 appropriately - (3.5) is equivalent to the classical design problem (3.1).

Steady-State Simulation

When we use independent replications as the variance estimation method, then steady-state simulation is similar to terminating simulation. We define an experiment as one run, and collect only one response from each run, being the mean of the observations made during the run. The SICOED design problem can then be used to determine an appropriate experimental design. Similarly, when we use regenerative simulation or batch means, we define an experiment as one (independent) sub-run, and collect the mean of the sub-run.

When we use a variance estimation method like spectral analysis, the SICOED design problem can still be applied, even though the design problem would appear to require the responses to be independent. To see why, we need to change the definition of an experiment, from one run to one observation within any run. Thus we now interpret n_i as the number of observations collected. In the development of the SICOED design problem we made the assumption that $\mathrm{Var}(\bar{y}(x_i)|n_i)$ is inversely proportional to n_i in (3.3), and this assumption was justified because the responses were assumed to be independent. The effect of this assumption is seen in the form of the cost function, which assumes that the cost of the experiment is inversely proportional to the value of σ_i^2 (note that $L(E^\sigma)$ is correctly based on the mean response variances).

In stochastic simulation there is usually serial correlation between observations. However the cost function of the SICOED approach is still (approximately) correct, provided we can make two assumptions about the stream of responses. Let $y_{i,j}$ be the response of the j^{th} experiment at design point i. A stream of responses is defined to be a weakly stationary stochastic process if $E[y_{i,j}]$ and $\mathrm{Cov}(y_{i,s}, y_{i,t})$ exist, and the relations $E[y_{i,s}] = E[y_{i,s+z}]$ and $\mathrm{Cov}(y_{i,s}, y_{i,t}) = \mathrm{Cov}(y_{i,s+z}, y_{i,t+z})$ are satisfied (Anderson (1971)). Then if we assume that the stream of responses at any design point is a weakly stationary stochastic process, and that

$$\sum_{z=-\infty}^{+\infty} \mathrm{Cov}(y_{i,j}, y_{i,j+z}) \leq \infty$$

then

$$\lim_{n_i \to \infty} n_i \mathrm{Var}(\bar{y}_i) = \sum_{z=-\infty}^{+\infty} \mathrm{Cov}(y_{i,j}, y_{i,j+z})$$

(Anderson (1971, Theorem 8.3.1.). This implies that asymptotically $\mathrm{Var}\left(\bar{y}(x_i) \middle| n_i\right)$ is inversely proportional to n_i, as required. In general, the assumptions made above would seem reasonable in the steady-state simulation context. In any case, any deviation from this assumption only impacts on the efficiency of the design through the cost function. Thus we are able to use the SICOED design approach for steady-state simulations where a variance estimation method like spectral analysis is used. *Once we know what the target value for the variance of the mean response is, we can choose any run-length and variance estimation method to achieve this target.*

In general, we are able to *interpret* n_i as either (i) the number of independent simulation runs performed (terminating or steady-state with independent replications), or (ii) the number of independent simulation sub-runs performed (steady-state with batch means or regenerative simulation), or (iii) the number of autocorrelated observations collected (steady-state with spectral analysis). Note the emphasis on the word 'interpret', since the interpretation of n_i is useful only for understanding the SICOED approach, as the variable n_i is not a part of the approach itself.

3.9. Advantages of the SICOED Approach

The SICOED approach overcomes all of the problems of the classical design property criteria and optimal design / sequential analysis combination approaches that were identified in section 4 of this chapter. As a result, the SICOED approach is suitable for implementation in simulation experimental design software.

In terms of flexibility, the experimenter is able to perform all the required experiments for a given design point before any experimentation is required at any one of the other design points. The design problem also allows any (convex) design region, provided it can be expressed as a series of constraints on the $\mathbf{x}_i$. These are then constraints in the design problem.

Unlike the classical design property criteria and optimal design approaches, the SICOED design problem explicitly allows, and takes advantage of, non-constant cost-per-experiment and variance functions. These functions must be set as part of the SICOED design problem (3.5). In general, it is unlikely that the experimenter will know the exact form and parameters of the cost-per-experiment function $c(\mathbf{x}_i)$ and the variance function $\upsilon^2(\mathbf{x}_i)$, as has been assumed thus far. However, the experimenter is likely to be able to provide estimates of these functions from previous experience with the simulation model, or a pilot experiment may be performed. These estimates are then used as part of the design problem. In general, by adding information about the cost-per-experiment and variance functions into the design problem, the actual experimental situation is modelled more closely, which can result in a more efficient design (see the examples in Chapter 5).

Unlike the classical approaches, the SICOED approach has the significant advantage that even when *estimates* of the cost-per-experiment function $c(\mathbf{x}_i)$ and/or variance function $\upsilon^2(\mathbf{x}_i)$ are used, the design criterion can still be accurately evaluated. This is because the sequential component of our approach ensures that the design criterion of (3.5) does not depend on those functions. As a result an optimal design found using (3.5) will ensure that the design criterion target L_0 is met regardless of the cost-per-experiment and variance functions used. Of course, the efficiency of the design does depend on these functions, since the cost function depends on the accuracy of the estimates $\hat{c}(\mathbf{x}_i)$ and $\hat{\upsilon}^2(\mathbf{x}_i)$ used.

The classical design problem (3.1) can also be modified to include non-constant variance and cost-per-experiment functions as seen in section 2. However, if poor estimates of these functions are used, the resulting design will not produce the design criterion value anticipated. On the other hand, the

SICOED design problem will always produce a design that meets the design criterion target, regardless of the accuracy of the cost-per-experiment and variance function estimates.

Since the design criterion can be accurately evaluated regardless of the accuracy of the cost-per-experiment and variance functions, the SICOED approach will always result in the expected distribution of information over the design region. This is in contrast to the optimal design / sequential analysis combination. An example illustrating this is shown in section 1 of Chapter 5.

For the SICOED approach, the same design problem can be used for both termination simulations and steady-state simulations. For steady-state simulations, no assumptions have been made about which method is used to estimate the variance of the mean response at each design point $\mathbf{x}_i$. Any one of the available methods such as Batch Means and Spectral Analysis, some of which can be significantly more efficient than independent replications, can be applied. If the responses are correlated, which is the case when n_i is defined as an individual observation, then the SICOED design problem (3.5) can still be applied provided an appropriate estimator for the variance of the mean response is used, and a reasonably large number of observations are collected. Also no assumptions are made regarding the way in which experiments are performed, so that the experimenter is free to choose whether this will be one long run or a number of shorter runs. For steady-state simulation of queueing models, the guidelines presented by Whitt (1989) can be used in this decision.

Two features of the SICOED approach ensure that it is suitable for use in simulation experimental design software. First, the experimental design is determined by solving an optimisation problem, as opposed to the arbitrary decisions required by the currently used classical design property approach. Second, by including a sequential element the overall size of the experiment is not assumed to be known. These features allow algorithms to be developed which perform the design phase of simulation experimentation, based on a small number of inputs. The required inputs are the specification of a design criterion, the metamodel to be fitted, the design criterion target, and design region.

Optionally, to improve the efficiency of the design, estimates for the cost-per-experiment and variance functions can be included. On the basis of these inputs, an experimental design can be determined, that (apart from the run-lengths in steady-state simulation) provides a complete specification of the experiments to be performed.

One drawback of the SICOED approach is that the experimenter is required to specify the design criterion target L_0. If the design criterion is chosen to represent the experimental objective, and some prior information is available (such as from a pilot experiment) then this choice should not be difficult. However, L_0 remains an absolute measure, and it would be preferable for experimenters to be able to specify a relative measure. This issue is discussed further in Chapter 6.

Lastly, although the emphasis has been on simulation, many of the advantages listed above also apply when the SICOED approach is used in a classical situation. In particular, there is a strong parallel between the problem of defining an experiment in the simulation and classical contexts. In the same way that defining an experiment as a single 'run' leads to a lack of attention to the choice of a suitable run-length, defining an experiment as growing a crop on a plot of land ignores the choice of the size of that plot.

3.10. Summary

In this chapter we first investigated two possible approaches, classical optimal design theory and sequential analysis, that might be more suitable for simulation than the methods used currently. Optimal design theory and sequential analysis are two bodies of theory that would appear to overcome some of the major problems facing automation of the experimental design process. We also proposed a third approach, being the combination of optimal design and sequential analysis. By combining these methods, both the selection of the design and the selection of the overall sample-size are able to be automated. However, as

Table 3.1. indicates, a number of aspects of all three approaches when used in the simulation context appear to be unsatisfactory. This includes a number of assumptions and inflexibilities, and some difficulties standing in the way of automation.

In the remainder of this chapter we have presented an alternative design approach, which we label as Semi-sequential Information Constrained Optimal Experimental Design, or SICOED. The focus of this approach is on ensuring that a desired amount of information is collected, rather than with staying within a budget. This approach overcomes most of the problems associated with current design methods, and is suitable for use in experimental design software. Although developed for use in a simulation context, the SICOED approach may also be useful in a number of classical contexts.

Figure 3.4. completes Figures 2.3. and 3.1. by adding the SICOED approach, and joining it to the contexts to which it can be applied.

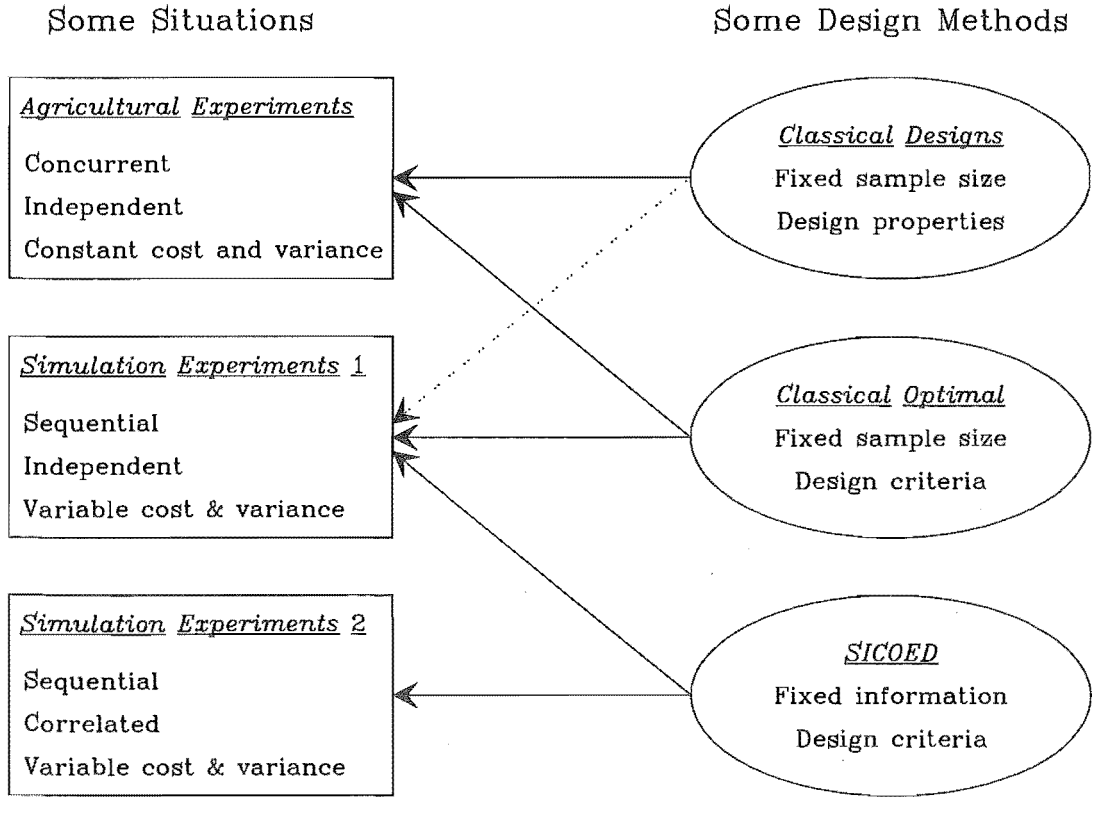


Figure 3.4. Some Experimental Situations and Design Methods (completed)

CHAPTER 4: CUSTOMISING AND SOLVING THE DESIGN PROBLEM

4.1. Introduction

In Chapter 3 a new approach to optimal experimental design was developed, which was labelled the SICOED approach. The design problem associated with this approach is a non-linear constrained optimisation problem, with the objective of minimising experimental costs while constraining the value of a design criterion.

To apply the SICOED approach, the experimenter needs to make a number of choices that determine the complete specification of the design problem, and then solve this problem to obtain the optimal design. In the first part of this chapter we identify and investigate the choices that the experimenter must make. In general the choice of factors, design region, metamodel form, and design criterion target depends strongly on the objective of the experiment and the specific experimental situation. More can be said about selection of the cost-per-experiment and variance functions, parameter estimation method, and design criterion. Sections 4.3. to 4.5. deal specifically with the alternatives available for those choices.

Due to the non-linear nature of many design criteria, finding the optimal design is usually not an easy task. Three basic methods can be employed: Algebraic solution methods, non-linear optimisation algorithms, and heuristic solution methods. In the second half of this chapter we investigate each of these methods in turn, as well as a number of important properties of the design problem.

4.2. Customising the SICOED Design Problem

In this section we consider the elements that make up the experimental design problem for the SICOED approach, and how the user may select those elements appropriately. The list below shows the seven choices that the experimenter must make to complete the design problem specification:

- The factors $x_1, ..., x_m$ assumed to influence the response y
- The design region $\mathcal{X}$
- The form of the metamodel relating the response and the factors
- The form and parameters of the cost-per-experiment $c(x_i)$ and variance $\upsilon^2(x_i)$ functions
- The parameter estimation method
- The design criterion L(E)
- The design criterion target L_0

The actual choices made should depend on the objectives of the experiment, and the experimental situation. For all of these choices, some information is available to aid the decision. To determine the appropriate factors, factor screening methods can be used, e.g. see Bettonvil and Kleijnen (1995) and the references there. The experimenter generally will have a good idea of which factor settings are to be investigated, and together with some prior information this leads to a design region. In some cases, such as when the design problem is part of an optimisation procedure, the design region may be determined by that procedure.

Except in very specific cases where a theoretical reason exists for a specific model, the choice of metamodel has generally been assumed to be restricted to the class of low order (linear or quadratic) polynomial models. One exception is Cheng and Kleijnen (1995) who provide a general model for queueing systems that are heavily loaded at some of the factor settings studied, which can not be adequately modelled using a simple polynomial model. However, the choice of model not only influences the accuracy of the fitted model, but it also forms the basis for the 'optimal' design problem, and hence determines a large part of the

efficiency of the design used. Usually some prior information is available about the experimental situation, and this can be used to help select the form of the metamodel. However, without a fully sequential experimental design procedure, the choice of metamodel will remain a difficult problem.

Similar comments apply to the choice of design criterion target. First, it is clearly important that the design criterion be chosen such that its values can easily be related to the experimental situation, so that an appropriate target can be set. Data from the validation and verification stages of the simulation model, or a pilot experiment, may be used here. Again, a fully sequential experimental design procedure would simplify such a choice enormously, as we would then be able to specify relative criterion targets (like 10%) rather than absolute targets as required by the current design problem.

Significantly more can be said about the choice of cost-per-experiment and variance functions, the parameter estimation method, and the design criterion itself. These are the subject of the next three sections.

4.3. The Variance and Cost-per-Experiment Functions

It was pointed out in section 9 of Chapter 3 that since the variance and cost-per-experiment functions are part of the objective function of the SICOED design problem, they impact only on the efficiency of the design. Hence it is not crucial that they be specified accurately, as the design criterion target will be reached regardless, although we would wish them to be as accurate as possible for efficiency reasons.

In general we know neither the form nor parameters of the variance function. One exception is the class of GI|G|m steady-state queueing models, for which good estimates of the asymptotic variance function for the mean number of waiting customers (and various other response variables for the M|M|1 queue) can be found in a paper by Whitt (1989). For example, the asymptotic variance

function for the mean time in the system of an M|M|1 queueing model is shown to be

$$\upsilon^2(\rho) = 2\rho^2\left(1+4\rho-4\rho^2+\rho^3\right)/(1-\rho)^4,$$

where ρ is the traffic intensity. This function will be used in the example in section 2 of Chapter 5. In other cases, an estimate of the variance function may be obtained from prior experiments or a pilot experiment.

Because the simulation literature has (implicitly) assumed that the cost per experiment is constant, very little has been published about this part of the design problem. However, the cost per experiment can vary substantially across the design region. For example, this is the case in discrete event simulation, when a factor determines the number of events that must be processed (see the example in section 3 of Chapter 5). Note that since the cost per experiment generally depends not only on the factors but also on random variability, then by 'cost per experiment' we will mean the expected cost.

We will now consider the cost-per-experiment functions for classical experiments, terminating simulations, and steady-state simulations separately. In all cases it is possible to use a constant function, if no estimate is available. However, although the design criterion target will be reached, a design based on a constant cost-per-experiment function may be very inefficient.

For many classical experiments the cost per experiment varies with the number of experiments performed at any design point. For example, this may be due to economies of scale, where the marginal cost of performing experiments at a particular combination of factor settings may drop as more experiments are performed. As it stands, the SICOED design problem assumes that the cost-per-experiment function depends only on the factor settings, and not the number of experiments performed at those factor settings. But if such a dependence does exist, then as long as we know the form of the dependence we can simply replace n_i with $\upsilon^2(x_i)/\sigma_i^2$ to obtain an estimate of the cost function. For example, if $c(x_i, n_i) = 2x_i/\sqrt{n_i}$, then the cost function becomes

$$\sum c(x_i, n_i) n_i = \sum 2x_i \sqrt{n_i} \approx \sum 2x_i \sqrt{\frac{\upsilon^2(x_i)}{\sigma_i^2}}.$$

Many of the comments below about simulation experiments also apply to classical experiments.

The cost function for terminating simulations is relatively simple. As part of the development of the SICOED design problem, we assumed that the cost function was given by $\sum c(x_i) n_i \approx \sum c(x_i) \upsilon^2(x_i)/\sigma_i^2$, i.e. linear in the number of experiments performed. For terminating simulations this is indeed the case, since each run at a particular design point uses the same expected amount of computer time. The cost-per-experiment function can then be estimated using a pilot experiment or data collected previously.

For steady-state simulation, the form of the cost function depends on the variance estimation method used, due to the initial transient period that is usually discarded. We consider three cases.

(a) When we use a variance estimation method such as Regenerative simulation, then no initial transient period is discarded. The (expected) cost function is then as for classical experiments, with no fixed cost component. If a pilot experiment is used to provide an estimate of the cost-per-experiment and variance functions, then the run-length of the pilot runs can be of any length.

(b) When we use a variance estimation method such as Spectral Analysis, we generally only have one warm-up period per run. Assume that we perform only one long run at each design point x_i, with an initial transient period of length w_i. Then the cost function is still $\sum c(x_i) n_i$, but now $n_i \neq \upsilon^2(x_i)/\sigma_i^2$ (remember that for steady-state simulation, n_i is defined as the number of observations collected). Instead, we have

$$n_i \approx \frac{\upsilon^2(x_i)}{\sigma_i^2} + w_i,$$

leading to the cost function

$$\sum\left(c(x_i)\frac{v^2(x_i)}{\sigma_i^2}+c(x_i)w_i\right).$$

Hence we have a single fixed cost for each design point at which an experiment is performed. If a pilot experiment is performed to obtain the data used to estimate the cost function, then any run-length and initial transient period length can be chosen for the pilot.

(c) When we use Independent Replications to determine the estimate of the variance of the mean response, then an experiment is defined as a single run. Since the run is of known length, then the cost of the initial transient period for each run is part of the cost per experiment. Like terminating simulations, there is no fixed cost component in the cost function. Instead, the total cost depends on the total number of runs performed. If a pilot experiment was performed, then the individual experiments in the pilot should be nearly identical (in terms of any parameters set, or other conditions) to the experiments specified by the subsequent design, to ensure that the estimates obtained are representative. When it is anticipated that the number of runs at any design point for the full experiment will be small, then such a pilot experiment (consisting of maybe as few as one run at any design point) may not be worthwhile. However in simulation we also control the size of each experiment. Thus we can set the length of the pilot runs to be shorter than the run-length used for the full experiment in order to allow a number of repetitions of the pilot design. To ensure that the relative cost-per-experiment function estimates obtained at different design points are representative, care must be taken to ensure that the ratio

$$\frac{\text{length of pilot run}}{\text{length of 'design' run}}$$

is the same for all design points. Otherwise if a pilot run at one design point is significantly longer, then the cost of that run will also be higher relative to the other design points. We believe that it might also be advisable to ensure that the ratio

$$\frac{\text{length of initial transient chosen}}{\text{length of run}}$$

for the pilot runs is the same as for the 'design' runs. This is so that the estimate of the variance function is based on the some proportion of the simulation run (i.e. that proportion that remains after the initial transient period is deleted) in both cases. However, the length of the initial transient period chosen for the pilot run is then known to be shorter than the actual length (in a statistical sense) of the transient period, which may also have an adverse impact on the estimate of the variance function.

Although we could estimate $c(\mathbf{x}_i)$ and $\upsilon^2(\mathbf{x}_i)$ separately, note that only the function $c(\mathbf{x}_i)\upsilon^2(\mathbf{x}_i)$ is required for the objective function of the SICOED design problem. A simple method, for when the cost per experiment is (or is approximated to be) only a function of the factor settings, is as follows. Let $T(\mathbf{x}_i,n)$ be the total cost of n experiments at $\mathbf{x}_i$, and $\text{var}(\overline{y}_i|n_i)$ the variance of the mean response obtained from those experiments. Since $c(\mathbf{x}_i) = T(\mathbf{x}_i,n_i)/n_i$, and $\upsilon^2(\mathbf{x}_i) \approx n_i \,\text{var}(\overline{y}_i|n_i)$ for large n_i, then an estimate of $c(\mathbf{x}_i)\upsilon^2(\mathbf{x}_i)$ is given by $T(\mathbf{x}_i,n_i)\text{var}(\overline{y}_i|n_i)$ for any given n_i. Clearly the larger n_i is, the better the estimate will be. Note that again, as in previous chapters, the exact meaning of n_i depends on the variance estimation technique used. For Spectral Analysis, n_i is the number of customers in a run, while for Independent Replications n_i is the number of runs.

4.4. The Estimators Used

The process of experimentation can be divided into three phases: (i) experimental design, (ii) carrying out the experiments, and (iii) analysing the data collected. The approach taken in this thesis is that in the experimental design phase we should attempt to model the analysis phase as closely as possible. For example, the design criterion chosen should reflect the objective of the experiment, which is generally related to the results of the analysis phase.

In this thesis we have assumed that the metamodel is linear in its parameters, and that those parameters are determined by the method of least squares. Thus far, we have assumed that Weighted Least Squares is used.

However, there are a number of variations of the least squares method, which are suitable for different situations. For independent response data, these are known as Ordinary Least Squares, Corrected Least Squares, Weighted Least Squares, and Estimated Weighted Least Squares. An important question then is: As part of the design methods investigated, which variation of the least squares method should be used? In general, it should be the least squares method that is used to analyse the experimental data once it has been collected. However, this is not always possible. In this section we investigate this question.

As before, we assume that n_i experiments are performed at r distinct design points x_i, leading to the responses $\{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$ and mean response

$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} \Big/ n_i .$$

The response variance is not assumed to be constant, but we still assume independence between responses (ensured by using independent random number streams for each simulation run). The vector of variances of the mean response at each design point will be denoted by the vector $\omega = [\mathrm{var}(\bar{y}_1) \quad \mathrm{var}(\bar{y}_2) \quad \ldots \quad \mathrm{var}(\bar{y}_r)]^T$. To fit the response model, four commonly used estimators are available, known as Ordinary Least Squares (OLS), Corrected Least Squares (CLS), Weighted Least Squares (WLS), and Estimated Weighted Least Squares (EWLS).

It is well known that the OLS estimator for the metamodel parameters is given (in vector form) by

$$\hat{\beta}^{OLS} = \left(\sum_{i=1}^{r} \frac{n_i}{\sigma^2} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1} \sum_{i=1}^{r} \mathbf{f}(x_i)\bar{y}_i .$$

This estimator is derived by assuming constant response variance (Draper and Smith (1981)). However, under mild technical assumptions this estimator is unbiased ($E\left[\hat{\beta}^{OLS}\right] = \beta$) and consistent ($\lim_{N\to\infty}\left(\hat{\beta}^{OLS}\right) = \beta$) for *any* response distribution

(Schmidt (1976, p65)). When the response variance is constant and equal to σ^2, then the covariance matrix of this estimator is

$$\Omega_{\hat{\beta}}^{OLS} = \mathrm{Cov}(\hat{\beta}) = \left(\sum_{i=1}^{r} \frac{n_i}{\sigma^2} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1}.$$

Unlike $\hat{\beta}^{OLS}$, $\Omega_{\hat{\beta}}^{OLS}$ is a biased estimator of the true covariance matrix when the response variance is not constant.

The second estimator, CLS, has the same parameter estimator as OLS. The difference lies in the parameter covariance matrix estimator, which does not assume that $\mathrm{var}(\bar{y}_i)$ is constant:

$$\Omega_{\hat{\beta}}^{CLS} = \left(\sum_{i=1}^{r} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1} \left(\sum_{i=1}^{r} \omega_i \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right) \left(\sum_{i=1}^{r} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1},$$

(Kleijnen and van Groenendaal (1992)) where ω_i is the i^{th} element of ω. In general ω is unknown, and only an estimate is available. But provided an unbiased estimator $\hat{\omega}$ of ω is used, then it can be shown that $\hat{\Omega}_{\hat{\beta}}^{CLS}$ is unbiased (Kleijnen and van Groenendaal (1992)).

However, of all the unbiased estimators the CLS estimator is not the minimum variance estimator when the response variance is not constant. That is given by the WLS estimator

$$\hat{\beta}^{WLS} = \left(\sum_{i=1}^{r} \frac{1}{\omega_i} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1} \sum_{i=1}^{r} \frac{1}{\omega_i} \mathbf{f}(x_i)\bar{y}_i,$$

with covariance matrix

$$\Omega_{\hat{\beta}}^{WLS} = \left(\sum_{i=1}^{r} \frac{1}{\omega_i} \mathbf{f}(x_i)\mathbf{f}^T(x_i) \right)^{-1}.$$

Again, in general ω is unknown and only an estimate is available.

But when we replace ω with an unbiased estimate $\hat{\omega}$, the resulting estimator, known as the EWLS estimator, may not have the same properties as

the WLS estimator. If the mean responses are normally distributed, $\hat{\beta}^{EWLS}$ is unbiased (Kleijnen and van Groenendaal (1992)). For other mean response distributions, Schmidt (1976) and van der Genugten (1983) show that if a consistent estimator of ω is used, and assuming a few limit conditions, the EWLS estimator is consistent and has the same asymptotic distribution as the WLS estimator. However, for small sample sizes EWLS may lead to a biased estimate of $\Omega_{\hat{\beta}}$.

Hence for situations with non-constant response variance and normally distributed responses, all of the least squares methods discussed above lead to unbiased estimators of β. However, for the estimator of $\Omega_{\hat{\beta}}$, OLS is biased, CLS is unbiased but inefficient, WLS requires ω to be known exactly, and EWLS is efficient but may be biased at small sample sizes. Thus for the analysis done after experimentation, we can remove OLS (since CLS is superior) and WLS (which requires unknown parameters to be specified) from consideration.

In Chapter 3, we assumed that the design criterion value was calculated using WLS, because the values of σ_i^2 specified by the design are not estimates but known exactly. However, we should ideally use the same estimation method in the design phase as we will use in the analysis phase. It is unlikely that the response variances are known exactly, and thus we cannot use WLS in the analysis phase.

Because the CLS estimators are unbiased, this variation of the least squares method would appear to be a good choice for use in our design approach. However there are a number of problems that result from the use of this estimator. First, because CLS is closely associated with OLS these two methods have similar properties. OLS assumes that the response variance is constant at each design point. The behaviour of CLS is also to favour designs that reflect this assumption. So when we evaluate designs using CLS, if there are significant differences between the values of $\text{var}(\bar{y}_i)$ then the design will not be very efficient. For example, in OLS if one of the $\text{var}(\bar{y}_i)$ values is large then this indicates to the estimation method that the common response variance may be larger than the remaining values of $\text{var}(\bar{y}_i)$ suggest. Consequently the design

criterion value increases to take account of this perceived increase in data variability. Yet this could be due simply to either a larger response variance at x_i, or because fewer observations are collected. CLS inherits similar behaviour. This effect, which leads to σ_i^2 values in the optimal design that are very similar, can be seen in the top-left graph of Figure 4.1. The graphs in that figure were drawn using a simple quadratic metamodel (3 parameter) with a single factor, and show the design criterion value as a function of the number of experiments or variance of mean response at a particular design points x_i. Notice that as $\text{var}(\bar{y}_i) \to \infty$, we would expect design point x_i to have less and less influence on the design criterion. However, Figure 4.1. shows that as $\text{var}(\bar{y}_i) \to \infty$, then $L(E) \to \infty$.

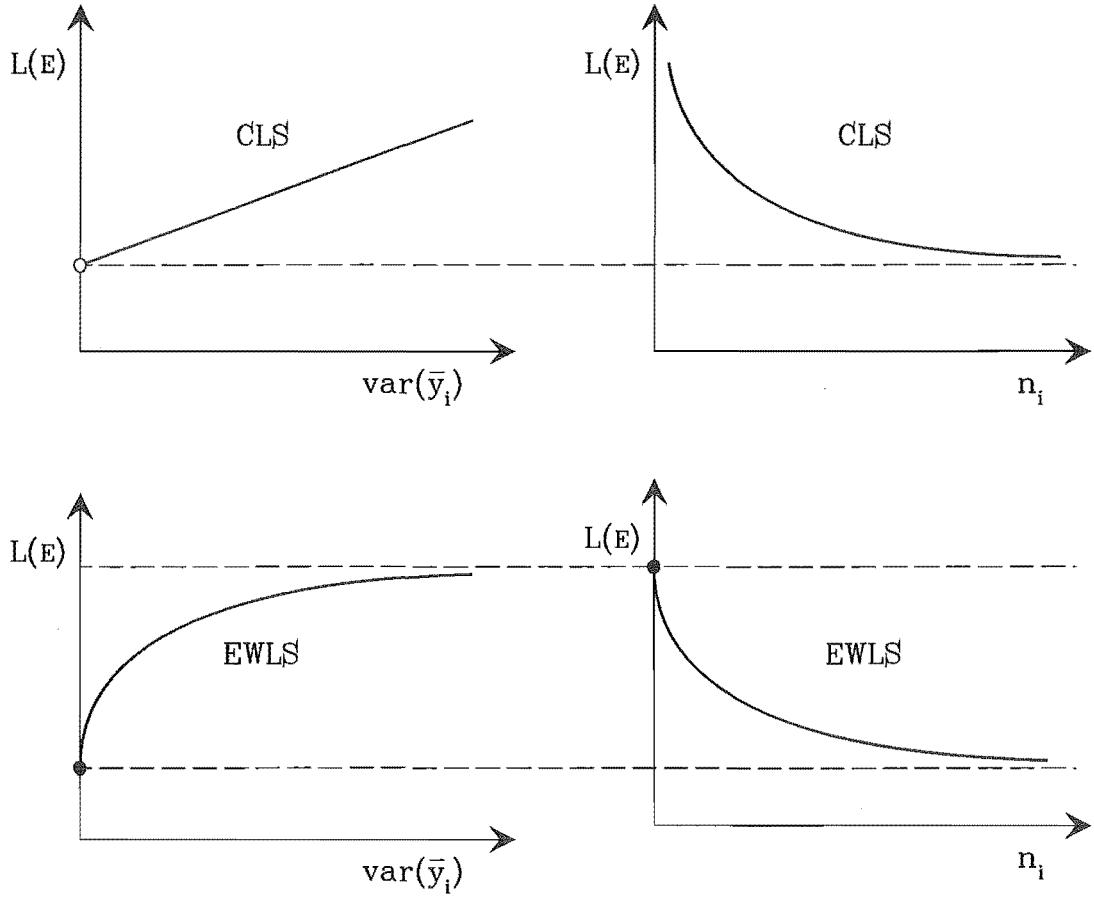


Figure 4.1. The behaviour of CLS and EWLS

Second, it would be very difficult to find the optimal design for any design problem that uses CLS. This is because any optimisation method (see section 6 onwards) needs to be able to remove a design point from consideration. However,

the nature of CLS means that this is not easily done. If we increase the value of σ_i^2, which would appear to be the most obvious way to reduce the influence of design point x_i, then we worsen the design criterion value. Decreasing the value of σ_i^2 leads to a higher cost design, and an improvement in the design criterion value. So the nature of CLS means that there is no value for σ_i^2 that will remove design point x_i from consideration. Any solution (design) would have $0 \leq \sigma_i^2 \leq \infty \ \forall i$. It would appear that time-consuming integer programming methods, or similar heuristics, might be required to solve the design problem if CLS is used.

On the other hand, EWLS does not have these problems. EWLS explicitly recognises that the response variance may be non-constant, and does not penalise large differences in the values of $\text{var}(\overline{y}_i)$. The values of $1/\text{var}(\overline{y}_i)$ are treated as weights, so that a design point with a large $\text{var}(\overline{y}_i)$ is effectively ignored, and has little or no impact on the design criterion value. This behaviour means that design points can easily be 'removed' from a design problem using EWLS by simply setting the appropriate σ_i^2 to be very large.

The EWLS estimator is also more efficient than the CLS estimator. Asymptotically the EWLS estimator is equivalent to the consistent and unbiased WLS estimator. Also as noted before, $\hat{\beta}^{EWLS}$ is unbiased for normally distributed data, and in most simulation situations this is a reasonable assumption as the 'data' are the means of run-length averages.

Unfortunately the EWLS estimators are not perfect. For small sample sizes, $\Omega_{\hat{\beta}}^{EWLS}$ may be biased. A number of papers have examined the extent of this bias through Monte Carlo simulations; see Kleijnen, Brent and Brouwers (1981), Deaton, Reynolds and Myers (1983), Nozari (1984), Kleijnen, Cremers and van Belle (1985), and Kleijnen (1992). Most report that between 10 and 25 response observations are required at each design point for the asymptotic distribution of EWLS to apply. For smaller sample sizes, $\Omega_{\hat{\beta}}^{EWLS}$ may substantially underestimate the true parameter covariance matrix.

However, EWLS still appears to be the best choice. In addition, in the design phase the EWLS and WLS estimators are equivalent, since the σ_i^2 values are known and not estimates. Although the EWLS estimators may be biased, the use of WLS in the design phase does closely 'model' the use of EWLS in the analysis phase.

Optimal experimental designs often have the property that the number of distinct design points is equal to the number of parameters in the metamodel. Such designs are known as saturated designs. If the number of distinct design points in the design (r) is equal to the number of metamodel parameters (p), then the CLS and WLS estimators are identical. This can be shown as follows. We use the property that for any square matrices A and B, $(AB)^{-1} = B^{-1}A^{-1}$. For simplicity of notation we define $X = \left[\mathbf{f}^T(\mathbf{x}_1) \quad \mathbf{f}^T(\mathbf{x}_2) \quad \cdots \quad \mathbf{f}^T(\mathbf{x}_r)\right]^T$, $\Omega_{\bar{y}} = \omega^T I$ where I is the identity matrix, and $\bar{y} = \left[\bar{y}_1 \quad \bar{y}_2 \quad \ldots \quad \bar{y}_r\right]^T$. If r = p, then X is square. Thus

$$\begin{aligned}
\hat{\beta}^{CLS} &= X^{-1}\left(X^T\right)^{-1}X^T\bar{y} \\
&= X^{-1}\bar{y} \\
\Omega_{\hat{\beta}}^{CLS} &= X^{-1}\left(X^T\right)^{-1}X^T\Omega_{\bar{y}}XX^{-1}\left(X^T\right)^{-1} \\
&= X^{-1}\Omega_{\bar{y}}\left(X^T\right)^{-1} \\
&= \left(X^T\Omega_{\bar{y}}^{-1}X\right)^{-1}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\beta}^{WLS} &= X^{-1}\Omega_{\bar{y}}\left(X^T\right)^{-1}X^T\Omega_{\bar{y}}^{-1}\bar{y} \\
&= X^{-1}\bar{y} \\
\Omega_{\hat{\beta}}^{WLS} &= \left(X^T\Omega_{\bar{y}}^{-1}X\right)^{-1}
\end{aligned}$$

Hence when r = p, $\hat{\beta}^{CLS} = \hat{\beta}^{WLS}$ and $\Omega_{\hat{\beta}}^{CLS} = \Omega_{\hat{\beta}}^{WLS}$. Since the CLS estimator remains unbiased when we replace $\Omega_{\bar{y}}$ with an unbiased estimate $\hat{\Omega}_{\bar{y}}$, then the EWLS estimator is also unbiased when r = p. This useful property is used in the examples in sections 1 and 2 of Chapter 5.

The second estimator that is used in the SICOED approach is the estimator of the variance of the mean response for each design point. Experiments are performed sequentially, and some stopping rule used to determine when the estimated variance s_i^2 is less than the required variance σ_i^2. The type of stopping rule seen most commonly in sequential procedures is to take an initial sample of size n_0, estimate $\text{var}(\bar{y}_i|n_0)$, and continue to perform another experiment until the estimate of $\text{var}(\bar{y}_i|n_i)$ is less than the required value. The usual variance estimator is used.

Kleijnen and van Groenendaal (1994) investigate this procedure, and show that the resulting estimator of $\text{var}(\bar{y}_i|n_i)$ is biased for small n_i. This can be seen as follows. If, after n_0 responses have been collected, the variance is overestimated, further experiments are performed (thus reducing the overestimate), while if the variance is underestimated the procedure may stop. Averaging over these outcomes shows that the mean response variance will be underestimated.

Unfortunately, alternative unbiased estimators do not appear to have been developed. We will investigate the effect of this bias on the SICOED approach in section 4 of Chapter 5.

4.5. The Design Criterion

The design criterion is an important component of the design problem, enabling the comparison of designs according to the amount of 'information' they provide. Traditionally the design criteria seen in the literature are the class of so-called alphabetic design criteria, such as D-, G-, E-, A-, and V-optimality. The most popular of these, D-optimality, is defined as the determinant of the inverse of the Fisher information matrix, $\left|M^{-1}(E)\right|$. Often the logarithm of this function is considered, as this leads to a convex function (this simplifies the process of finding the optimal design; see section 8 of this chapter).

The choice of the design criterion is an important part of the design process, and can have a substantial impact on the optimal design. For example, the classical D-optimal design for a quadratic function in one variable (ranging from -1 to 1) is to perform $1/3$ of the experiments at the points -1, 0 and 1. Atkinson and

Donev (1992, p110) show that if we restrict our attention to the D-optimal design for the quadratic coefficient, i.e. we consider only a part of the information matrix, then the optimal design is the weights ($^1/_4$, $^1/_2$, $^1/_4$) at the same design points. Just by considering a subset of the parameters, the design changes significantly.

However, the justification behind most of the alphabetic optimality criteria is simply that they have certain statistical properties which are considered to be desirable. For example, the value of $|M^{-1}(E)|$ (D-optimality) is related to the volume of the ellipsoidal confidence region for the parameters of the fitted model. Similarly, A- and E-optimality are related to the eigenvalues of M(E), which are also connected with the confidence region.

As mentioned before in Chapter 3, we believe that the selection of a design criterion should be based on the objective of the experiment, rather than general statistical properties. There are two reasons for this. First, an experimental design problem is used in order to obtain a design that will achieve our objectives. This includes both cost and information objectives. However if we use a design criterion that does not specifically reflect our information objective, then we are unable to ensure that that objective is realised. Second, the value of a criterion based on general statistical properties may be difficult, if not impossible, to interpret. For such criteria we are likely to have little idea of the range of values that would be acceptable. On the other hand, the experimenter's own information objective is closely related to the actual situation. Criteria based on such an objective will generally have more easily interpretable values.

In the remainder of this section, a number of specific design criteria are discussed, as well as various types of criteria.

In general, design criteria can be classified according to whether or not they are evaluated at a point in the design region. For example, the variance of the fitted response depends on the factor settings at which it is evaluated, but the variance of a fitted model parameter does not. First we consider the former type of criterion.

Let $h(\mathbf{x})$ be some function of the parameters of the correct metamodel evaluated at $\mathbf{x}$, and $\hat{h}(\mathbf{x})$ the estimate of $h(\mathbf{x})$ obtained from the fitted model (which may be a different model). Then a general quadratic loss design criterion, based on Box and Draper's (1959) J criterion, is

$$J' = \int_{\mathcal{I}} w(\mathbf{x}) E\left[h(\mathbf{x}) - \hat{h}(\mathbf{x})\right]^2 d\mathbf{x} \Big/ \int_{\mathcal{I}} d\mathbf{x},$$

where $w(\mathbf{x})$ is the value of a weight, or "relative importance", function for point $\mathbf{x}$. A suitable weight function would have $\int_{\mathcal{I}} w(\mathbf{x})d\mathbf{x} = 1$, $w(\mathbf{x}) \geq 0 \ \forall \mathbf{x}$. Note that the values of the weight function determine the shape and size of the region of interest $\mathcal{R}_i$. The idea of weighting the design space is rarely seen in the general experimental design literature, but this idea has surfaced in the Bayesian literature, e.g. Chaloner (1984).

Alternatively, we could use the weighted maximum rather than the weighted average:

$$J'' = \max_{\mathbf{x}} \ w(\mathbf{x}) E\left[h(\mathbf{x}) - \hat{h}(\mathbf{x})\right]^2,$$

and use a weight function that has $\max(w(\mathbf{x})) = 1$, $w(\mathbf{x}) \geq 0 \ \forall \mathbf{x}$.

Both J' and J'' allow the difference between $h(\mathbf{x})$ and $\hat{h}(\mathbf{x})$ to be due to a combination of sampling (or variance) error and bias (model misspecification) error. However, as noted in section 4 of Chapter 1 the inclusion of bias error in the design criterion generally leads to a design that is dependent on the unknown parameters of the correct model (Box and Draper (1959)). To obtain a design, 'typical' values of these parameters must be considered. Some advances in this area have been made in the simulation design literature, e.g. see Donohue, Houck and Myers (1992,1993a), but the procedures required to find such designs are still fairly complex. A simpler approach is one borrowed from the Bayesian literature. Let $\{L_1(E), L_2(E), ..., L_d(E)\}$ be a set of d design criteria, each one associated with a particular model for which we believe the probability that it is the correct model to be greater than zero. Then we replace the single constraint $L_1(E) \leq L_0$ with either

$$\sum_{i=1}^{d} p_i L_i(E) \le L_0,$$

where p_i is the prior probability (in a Bayesian sense) that the model corresponding to $L_i(E)$ is the true model, or

$$\begin{aligned} L_1(E) &\le L_0 \\ L_2(E) &\le L_0 \\ &\vdots \\ L_d(E) &\le L_0, \end{aligned}$$

depending on whether the experimenter is happy with $E[L(E)] \le L_0$. From now on we will assume that either there is no bias error, or that one of the above constraint sets is used.

The most commonly cited objective in the RSM literature is to estimate the *mean response* over an interval. Hence we set $h(x) = y(x)$. If we assume that the model to be fitted has been correctly specified, then from linear regression theory we obtain

$$\begin{aligned} E\left[y(x) - \hat{y}(x)\right]^2 &= \mathrm{Var}(\hat{y}(x)) \\ &= \mathrm{Var}(f^T(x)\hat{\beta}) \\ &= f^T(x)\left(\sum_{i=1}^{r} \frac{1}{\sigma_i^2} f(x_i) f^T(x_i) dx\right)^{-1} f(x) \\ &= f^T(x) M^{-1}(E^{\sigma}) f(x). \end{aligned} \tag{4.1}$$

We now have a mean response criterion by substituting (4.1) into either one of the general design criteria above. Note that for this 'variance of the mean response' criterion, setting a suitable value of L_0 should not be difficult, as the criterion has a clear interpretation.

Another common objective in RSM is to estimate the *slope* at a point or over a region. This is particularly relevant when the current experiment is part of an optimisation process such as steepest ascent, which requires the first partial derivatives of the model. However, estimating the slope consists of estimating the

partial derivatives with respect to each of the factors, since a slope has direction. Hence except in the case of a single factor, a decision also needs to be made about which direction the experimenter is interested in.

For single factor design problems, the variance of the least squares estimator of an arbitrary linear function of the regression coefficients, $\mathbf{c}^T\beta$ (where c is an arbitrary vector) is

$$\begin{aligned} \mathrm{Var}(\mathbf{c}^T\hat{\beta}) &= \mathbf{c}^T\mathrm{Var}(\hat{\beta})\mathbf{c} \\ &= \mathbf{c}^T M^{-1}(E^s)\mathbf{c}, \end{aligned}$$

(Murty and Studden (1972)). In the alphabetic optimality literature this is labelled c-optimality. Now if we let $\mathbf{c} = (0, 1, 2x, 3x^2, ..., kx^{k-1})^T$, then we get the variance of the slope of the polynomial model $y = b_0 + b_1x + b_2x^2 + ... + b_kx^k$, evaluated at the point x. For a general linear model in one factor, let $\mathbf{c}(x)$ be defined by

$$\mathbf{c}(x) = \begin{bmatrix} \partial f_1(x)/\partial x \\ \partial f_2(x)/\partial x \\ \vdots \\ \partial f_p(x)/\partial x \end{bmatrix},$$

where $f_i(x)$ is the i^{th} element of $\mathbf{f}(x)$. The design criterion for the objective of estimating the slope of the model then has

$$E\left[\mathbf{c}(x)^T\beta - \mathbf{c}(x)^T\hat{\beta}\right]^2 = \mathbf{c}(x)^T M^{-1}(E^\sigma)\mathbf{c}(x),$$

which is very similar to the mean response criterion.

Often there is more than one factor in the experiment. Park (1987) investigated the necessary conditions for slope-rotatable, multiple factor, variance-only slope designs. The following uses some of the results from that paper. Let the vector of first partial derivatives $\mathbf{g}(x)$ be defined by

$$\mathbf{g}(x) = \begin{bmatrix} \partial\left(\mathbf{f}^T(x)\beta\right)/\partial x_1 \\ \partial\left(\mathbf{f}^T(x)\beta\right)/\partial x_2 \\ \vdots \\ \partial\left(\mathbf{f}^T(x)\beta\right)/\partial x_m \end{bmatrix} = D_x\beta,$$

where the subscript on the matrix D is a reminder that it depends on the vector of factor settings. Assume that the experimenter is interested in the slope in a direction specified by the unit vector

$$\mathbf{k}^T = (k_1, k_2, \ldots, k_m), \qquad \sum_i k_i^2 = 1.$$

The variance of the slope in the direction of $\mathbf{k}$ is

$$\begin{aligned} \mathrm{Var}\left(\mathbf{k}^T D_x \hat{\beta}\right) &= \mathbf{k}^T D_x \,\mathrm{var}(\hat{\beta}) D_x^T \mathbf{k} \\ &= \mathbf{k}^T D_x M^{-1}(E^\sigma) D_x^T \mathbf{k}, \end{aligned}$$

(Park (1987)) and so the design criterion for estimating the variance of the slope in the direction of k has

$$E\left\{\mathbf{k}^T \mathbf{g}(\mathbf{x}) - \mathbf{k}^T \hat{\mathbf{g}}(\mathbf{x})\right\}^2 = \mathbf{k}^T D_x M^{-1}(E^\sigma) D_x^T \mathbf{k}.$$

Park also shows that the average slope variance over all directions is

$$\underset{\mathbf{k}}{\mathrm{avg}}\left[\mathrm{Var}\left(\hat{\mathbf{g}}(\mathbf{x})\right)\right] = \frac{1}{p}\mathrm{tr}\left[D_x M^{-1}(E^\sigma) D_x^T\right],$$

which can also be used as part of the design criterion.

However, as mentioned before, the experimenter is often interested in the direction of steepest ascent, rather than the slope in a given direction or all directions. The direction of steepest ascent is given by $\mathbf{g}(\mathbf{x}) = D_x\beta$, and the variance of this direction at $\mathbf{x}$ is

$$\mathrm{Var}(\mathbf{g}(\mathbf{x})) = D_x M^{-1}(E^\sigma) D_x^T.$$

Note that this is a matrix, the diagonal entries of which are the variances of the axial direction components. We can then take an average of these:

$$E\left\{\mathbf{g}(\mathbf{x}) - \hat{\mathbf{g}}(\mathbf{x})\right\} = \mathrm{tr}\left[D_x M^{-1}(E^\sigma) D_x^T\right],$$

(Myers and Lahoda (1975) investigate a continuous version, including bias considerations). Alternatively, we can have a criterion for each of the diagonal entries separately:

$$E\{g(x)-\hat{g}(x)\}_i^2 = \mathbf{d}_i M^{-1}(E^\sigma)\mathbf{d}_i^T, \qquad i=1,\ldots,m,$$

where $\mathbf{d}_i^T$ is the i^{th} row of D_x.

One problem with slope criteria is setting an appropriate value for L_0. In general, the experimenter will not know what the value of the slope is, and hence may find it difficult to determine an appropriate variance limit for it. This problem may be overcome by using a pilot experiment, which would provide an estimate of the slope, or by using sequential design procedures as discussed in Chapter 6.

Apart from the general criteria that depend on the factor settings at which they are evaluated, there is another set of criteria that are a function of the fitted parameters only. One example is provided by Cheng and Kleijnen (1995). They consider the metamodel

$$y_i = (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2)/(1-x_i) + \varepsilon_i,$$

and assume that the objective of the simulation is to estimate $p = \gamma_0 + \gamma_1 + \gamma_2$, leading to the design criterion $L(E) = Var(\hat{p})$. However such criteria are less suitable for use in our approach due to the difficulty of finding an appropriate target L_0. Generally, an acceptable value for $Var(\hat{p})$ would depend on the value of p itself, which is unknown. Again a pilot experiment can be used here to overcome this problem to some extent, although a fully sequential procedure would allow design criteria like $Var(\hat{p})/p$ to be used, for which a suitable target is more easily determined.

In some cases, the experimenter would be interested in not only finding a functional relationship for the mean of the response, but also for the variance of the response. This is a problem that has received little attention in the design literature, mainly because of the assumption of constant variance. Atkinson and

Cook (1993) investigate this problem, and provide theory for finding D-optimal designs for a particular class of variance functions. However the resulting designs, like the designs that take account of bias error, clearly depend on the unknown parameters of the variance function. This is because the variance function, which we are trying to estimate, is a crucial component of the design problem.

4.6. Solving the SICOED Design Problem

In the remainder of this chapter we briefly consider three approaches to solving the SICOED design problem - algebraic solution, non-linear optimisation methods, and heuristic methods. Each approach is evaluated in terms of speed, simplicity, and suitability for implementation as part of experimental design software.

The design problems for the classical and SICOED approaches are similar, and as noted in Chapter 3 the same design will be optimal for both design problems provided N and L_0 are set appropriately and the remaining assumptions and parameters are the same. The main differences between the design problems, besides simple restrictions on the variables, are that (i) the design variables are different, being $(\mathbf{x}_i, p_i)$ or $(\mathbf{x}_i, n_i)$ for the classical approach, and $(\mathbf{x}_i, \sigma_i^2)$ for the SICOED approach, and (ii) the classical design problem consists only of an objective function (the design criterion) while the design problem for the SICOED approach consists of an objective function (experimental cost) and a constraint (design criterion). As might be expected, the SICOED design problem has many of the properties of the classical design problem. Thus the solution methods considered in this chapter, particularly the heuristic method, are generally modified versions of methods found in the extensive literature for the construction of classical optimal experimental designs.

The design criterion is the most complex part of any design problem, and is usually highly non-linear. Hence most properties of a design problem depend on the design criterion chosen. Different design criteria result in different properties.

For example the behaviour of the D-optimality criterion, being the *determinant* of the information matrix, will be quite different to the G-optimality criterion, being the *trace* of the information matrix.

In the remainder of this chapter we will assume that the experiments are to be compared using the criterion $L\left(M^{-1}(E^{\sigma})\right)$, which associates a scalar with every design E^{σ}. For reference $M(E^{\sigma})$ is again defined as

$$M(E^{\sigma}) = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} \mathbf{f}(x_i)\mathbf{f}^T(x_i). \tag{4.2}$$

We assume that $L(\cdot)$ is a linear criterion (Fedorov (1972)), so that

$$L(A+B) = L(A)+L(B), \tag{4.3}$$

$$L(cA) = cL(A), \tag{4.4}$$

for all matrices A and B and scalars c. We also assume that

$$L(C) \geq 0, \tag{4.5}$$

for all positive semi definite matrices C. Note that for any classical design E the information matrix M(E) is symmetric and positive semi-definite (Fedorov (1972, Theorem 2.1.2)), and this property clearly also applies to SICOED designs E^{σ}.

Looking at Chapter 3, all of the various criteria discussed there are linear criteria. For example for the average variance of the fitted response, L acts on $M^{-1}(E^{\sigma})$ as

$$\int_{\mathcal{I}} \mathbf{f}^T(x)M^{-1}(E^{\sigma})\mathbf{f}(x)dx.$$

This criterion satisfies (4.3)-(4.5):

$$\int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})(\mathbf{A}+\mathbf{B})\mathbf{f}(\mathbf{x})d\mathbf{x} = \int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})\mathbf{A}\mathbf{f}(\mathbf{x})d\mathbf{x} + \int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})\mathbf{B}\mathbf{f}(\mathbf{x})d\mathbf{x},$$

$$\int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})c\mathbf{A}\mathbf{f}(\mathbf{x})d\mathbf{x} = c\int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})\mathbf{A}\mathbf{f}(\mathbf{x})d\mathbf{x},$$

$$\int_{\mathcal{X}} \mathbf{f}^T(\mathbf{x})\mathbf{C}\mathbf{f}(\mathbf{x})d\mathbf{x} \geq 0,$$

the last of these resulting directly from the definition of a positive semi definite matrix, being that C is symmetric and $\mathbf{a}^T\mathbf{C}\mathbf{a} \geq 0$ for all real vectors $\mathbf{a}$.

In fact the class of criteria that satisfies (4.3)-(4.5) includes a wide range of criteria that are related to the variability of the parameters, such as any criterion defined as the variance of $\mathbf{l}^T\hat{\boldsymbol{\beta}}$ for any real vector $\mathbf{l}$. As discussed in section 5 of this chapter, the values of such criteria are generally easily interpreted, and are thus very suitable for the SICOED approach.

4.7. Algebraic Solution Method

Probably due to the absence of sufficient computer power, the method used to solve the classical design problem in early papers on experimental design was to derive the solution algebraically (e.g. see Kiefer and Wolfowitz (1959), Box and Draper (1959)). In the one paper on optimal experimental design for simulation, Cheng and Kleijnen (1995) also use this approach. The advantages of this approach are that no convexity results (see the next section) are needed to ensure that the optimal design is globally optimal, and that if there are multiple global optimal designs then they can usually be found with little extra effort. In addition, this approach usually provides some insight into the design problem.

However the form of the algebraic solution depends heavily on the assumptions and choices that determine the design problem under consideration. In particular, the solution generally depends on the form of the metamodel, the design criterion, the cost-per-experiment and variance functions, and the shape of the design region. Different assumptions and choices will in most cases lead to a different form for the solution, and thus require the algebraic solution to be derived again. Also, this derivation may sometimes be difficult or impossible, and generally requires a high level of mathematical knowledge.

Algebraic solution approaches are complicated, and it would be difficult, if not impossible, to incorporate such approaches into easy-to-use experimental design software. On the other hand, this is not generally true for numerical solution methods.

4.8. Convexity, and the Modified SICOED Design Problem

Before we consider specific numerical optimisation methods, in this section we consider the convexity properties of the SICOED design problem. These properties impact on both the ability of numerical optimisation methods to determine the optimal design, and the form of the optimal design itself. For the classical optimal design problem, corresponding properties can be found in a number of texts on classical optimal design theory, such as Fedorov (1972), Pazman (1986), Pukelsheim (1993), and Silvey (1980).

Assume we have a design E_L^σ found by applying a numerical optimisation method to the SICOED design problem. E_L^σ is a *local* optimal design, meaning that no other design in a small neighbourhood around E_L^σ, in $(\mathbf{x}_i, \sigma_i^2)$ space, satisfies the design problem constraints and has a lower experimental cost. However, in general this does not imply that E_L^σ is also the *global* optimal design, which minimises the experimental cost over the entire set of designs that satisfy the design problem constraints. In order to ensure that E_L^σ is the global optimal design, the design problem must have a number of properties.

In particular, a local optimal solution to the problem of minimising a convex function over a convex set is also a global optimal solution (Bazaraa and Shetty (1979, Theorem 3.4.2)). Hence we must show that (i) the set of designs that satisfy the design problem constraints is a convex set, and (ii) that the experimental cost function is a convex function over this set. Since the set $S_\alpha = \{\mathbf{d} \in S : f(\mathbf{d}) \leq \alpha\}$ for any convex set S, vector of variables $\mathbf{d}$ and convex function $f(\cdot)$, is a convex set (Bazaraa and Shetty (1979, Lemma 3.1.2)), then property (i) requires that the design criterion $L(E^\sigma)$ be a convex function of the variables $\mathbf{x}_i$ and σ_i^2.

However the SICOED design problem generally does not have these properties. Due to the division by σ_i^2, the cost function of the SICOED design problem,

$$\sum_{i=1}^{r} \frac{c(\mathbf{x}_i)\upsilon^2(\mathbf{x}_i)}{\sigma_i^2},$$

is generally not a convex function over $\mathcal{X}$ even if $c(\mathbf{x}_i)\upsilon^2(\mathbf{x}_i)$ is a convex function over $\mathcal{X}$. Further, the design criterion is also generally not a convex function, and in fact has several asymptotes. The following two examples illustrate this.

Example 1: This example has one factor, ranging from 0 to 1, and two design points x_1 and x_2. The design criterion is the average variance of the mean response over the design region (see section 5 of this chapter). We set $\sigma_1^2 = \sigma_2^2 = 2$. Figure 4.2. shows the value of the design criterion, plotted as a function of the *position* of the two design points x_1 and x_2.

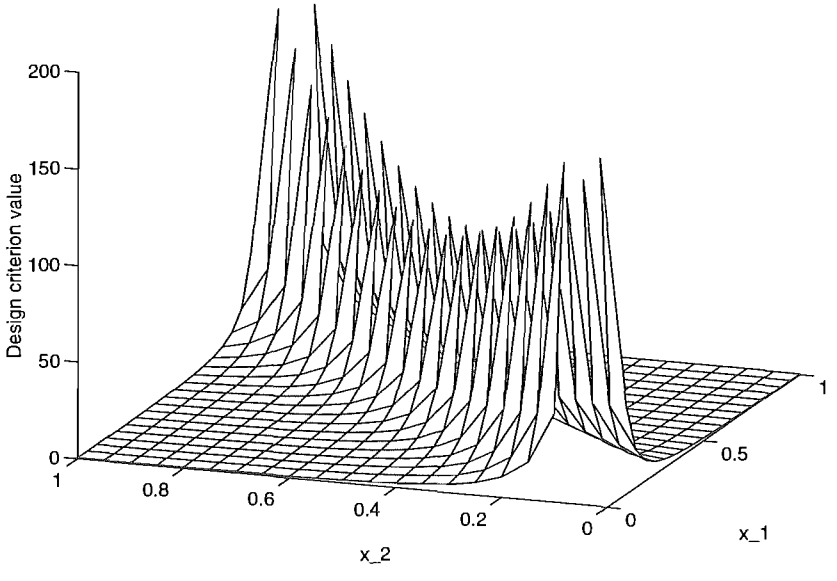


Figure 4.2. Non-convexity of the design criterion function - Example 1

When $x_1 = x_2$, there is only one distinct design point instead of two, and thus the value of the design criterion tends to infinity along the line $x_1 = x_2$. Without

restricting the design region to prevent this, one or more occurrences of such asymptotic behaviour will be present in (x_i, σ_i^2) space for almost all design criteria functions. This is because regardless of the number of design points (in relation to the number of metamodel parameters), it is always possible to find one or more settings for the x_i so that the information matrix M is singular.

The explanation for the apparent symmetry in the graph is that by re-arranging the subscripts on the variables for any particular design (i.e. swap x_1 and x_2), then we have the 'same' solution (the design criterion value will be the same) but we are in a different part of the solution space. In fact, in general if there are r distinct design points, then there are r! permutations of the design points of any optimal design. Each permutation of the design lies in a different part of the (x_i, σ_i^2) space, but has the same design criterion value. This implies that the design criterion is generally a non-convex function.

For this example, the simple restriction $x_1 < x_2$ (or equivalently, $x_1 > x_2$) will remove both the asymptote and the 'mirror image', and similar constraints can be added for any design problem with a single factor. However, finding such constraints for design problems with multiple factors (and non-cuboidal design regions) may not be easy, or even possible.

Example 2: This example has two factors, each ranging from 0 to 1. A standard factorial design is used, where the design point (0, 0.5) has been moved to (0, 0.35) and the design point (1, 0.5) has been moved to (0.65, 0.5), see Figure 4.3.

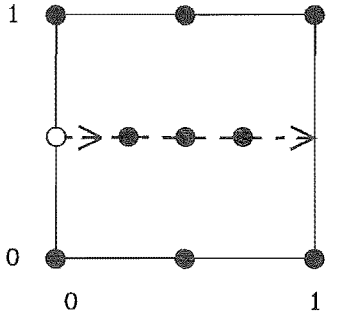


Figure 4.3. Design for Example 2

All the values of σ_i^2 are 0.02, except at the centre point where it is 0.01. We then add a 10th design point at (0.5, k), where k ranges from 0 to 1. Figure 4.4. shows a plot of the value of the design criterion (which is the same as for example 1) versus k.

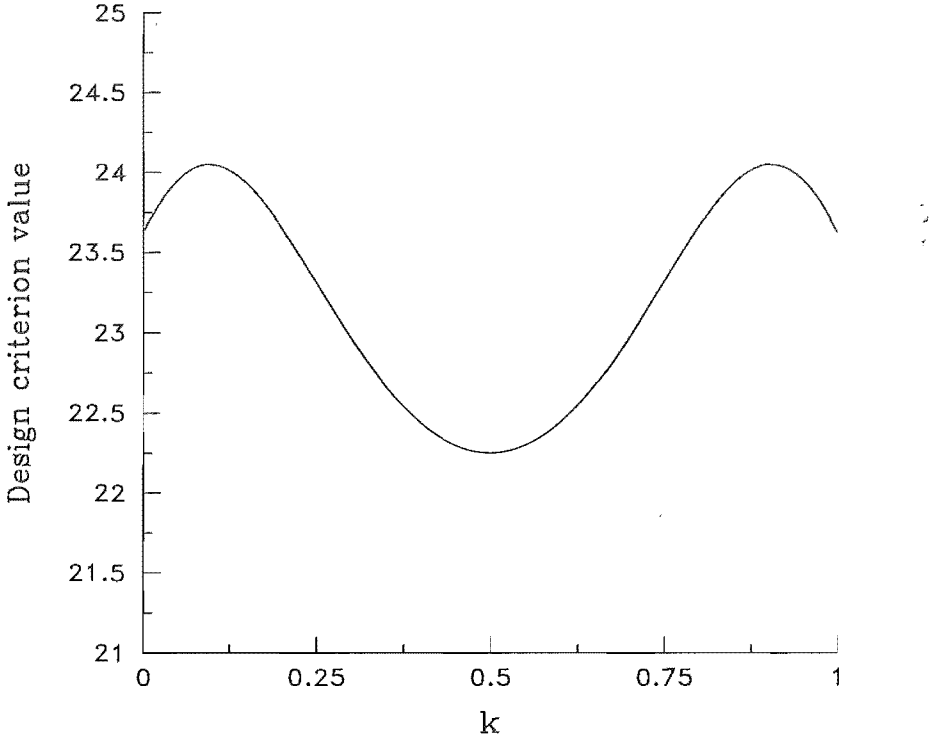


Figure 4.4. Non-convexity of the design criterion function - Example 2

Note that at k = 0.35, k = 0.5, and k = 0.65 the 10th design point coincides with one of the factorial design points. In this case, the design criterion is certainly not convex, and it seems reasonable to conclude that in general the design criterion is likely to be a non-convex function.

However if we make two modifications to the SICOED design problem then we can derive a number of useful properties for the modified design problem. The first modification is to remove the design point variables $\mathbf{x}_i$ from the problem by assuming that $\mathcal{X}$ consists of a finite number, t, of design points. For example, we may place a grid of equally spaced design points over $\mathcal{X}$. We will

show that x_i is fixed by using the notation $\dot{x}_i$ The second modification is to re-define the design in terms of the variables $\Sigma_i = 1/\sigma_i^2$, so that the optimisation problem consists of optimising the values of Σ_i rather than σ_i^2. We define the design $E^{1/\sigma}$ to be the collection of pairs

$$(\dot{x}_1, \Sigma_1), (\dot{x}_2, \Sigma_2), \ldots, (\dot{x}_t, \Sigma_t).$$

Once the optimal design $E^{1/\sigma}$ has been found, it is straightforward to convert this to the design E^{σ}.

These modifications ensure that the modified SICOED design problem

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{r} c(\dot{x}_i)\upsilon^2(\dot{x}_i)\Sigma_i \\ \text{s.t.} \quad & L(E^{1/\sigma}) \le L_0 \\ & \Sigma_i \ge 0 \quad \forall i \end{aligned} \tag{4.6}$$

has a number of desirable properties as shown by the theorems below.

First, let the design $E_1^{1/\sigma}$ be the collection of pairs $\{(\dot{x}_{1,1}, \Sigma_{1,1}), (\dot{x}_{2,1}, \Sigma_{2,1}), \ldots, (\dot{x}_{t,1}, \Sigma_{t,1})\}$, and the design $E_2^{1/\sigma}$ the collection of pairs $\{(\dot{x}_{1,2}, \Sigma_{1,2}), (\dot{x}_{2,2}, \Sigma_{2,2}), \ldots, (\dot{x}_{t,2}, \Sigma_{t,2})\}$. Since the $\dot{x}_i$ are fixed, the design $\alpha E_1^{1/\sigma} + (1-\alpha)E_2^{1/\sigma}$, where $0 \le \alpha \le 1$, is then the collection of pairs

$$\{(\dot{x}_{1,1}, \alpha\Sigma_{1,1}), (\dot{x}_{2,1}, \alpha\Sigma_{2,1}), \ldots, (\dot{x}_{t,1}, \alpha\Sigma_{t,1}), (\dot{x}_{1,2}, (1-\alpha)\Sigma_{1,2}), (\dot{x}_{2,2}(1-\alpha(\Sigma_{2,2}), \ldots, (\dot{x}_{t,2}, (1-\alpha)\Sigma_{t,2})\}.$$

We will use this in the next theorem.

Theorem 4.1.

$L\left(M^{-1}(E^{1/\sigma})\right)$ is a convex function of Σ_i, i = 1,...,t.

Proof: The set of all designs $E^{1/\sigma}$ is a convex set, as the Σ_i are unbounded. It remains to show that the design criterion is a convex function, by showing that

$$L\left(M^{-1}\left(\alpha E_1^{1/\sigma} + (1-\alpha)E_2^{1/\sigma}\right)\right) \le \alpha L\left(M^{-1}\left(E_1^{1/\sigma}\right)\right) + (1-\alpha)L\left(M^{-1}\left(E_2^{1/\sigma}\right)\right).$$

Using (4.2) we obtain

$$M\left(\alpha E_1^{1/\sigma} + (1-\alpha)E_1^{1/\sigma}\right) = \alpha M(E_1^{1/\sigma}) + (1-\alpha)M(E_2^{1/\sigma}).$$

Since the difference

$$\alpha A^{-1} + (1-\alpha)B^{-1} - (\alpha A + (1-\alpha)B)^{-1}$$

is a positive definite matrix for any positive definite matrices A and B (Fedorov (1972, Theorem 1.1.12.), then using (4.3) and (4.5) we obtain

$$\begin{aligned} L\left(M^{-1}\left(\alpha E_1^{1/\sigma} + (1-\alpha)E_2^{1/\sigma}\right)\right) &\le L\left(\alpha M^{-1}\left(E_1^{1/\sigma}\right) + (1-\alpha)M^{-1}\left(E_2^{1/\sigma}\right)\right) \\ &= \alpha L\left(M^{-1}\left(E_1^{1/\sigma}\right)\right) + (1-\alpha)L\left(M^{-1}\left(E_2^{1/\sigma}\right)\right) \end{aligned}$$

as required.◊

We define the set of all *feasible* designs, or designs that meet the constraints of the modified SICOED design problem (4.6), as

$$\{E^{1/\sigma} : \Sigma_i \ge 0 \ (i = 1,..,t),\ L(E^{1/\sigma}) \le L_0\}.$$

Hence for a feasible design, (i) the target variances σ_i^2 are positive, and (ii) the design criterion value lies at or below the acceptable limit L_0.

Theorem 4.2.

The set of all *feasible* designs for the modified SICOED design problem (4.6),

$$\{E^{1/\sigma} : \Sigma_i \ge 0 \ (i = 1,..,t),\ L(E^{1/\sigma}) \le L_0\}.$$

is a convex set.

Proof: The set $S_\alpha = \{\mathbf{d} \in S : f(\mathbf{d}) \le \alpha\}$ for any convex set S, vector of variables $\mathbf{d}$ and convex function $f(\cdot)$, is a convex set (Bazaraa and Shetty (1979, Lemma 3.1.2)). Since the set of all designs (being unbounded) is a convex set, $-\Sigma_i$ is a

convex function, and $L(E^{1/\sigma})$ is a convex function by Theorem 4.2., then the set of feasible designs is convex.◊

Lemma

A local optimal solution to the modified SICOED design problem (4.6) is also a global optimal solution.

Proof: In general, a local optimal solution to the problem of minimising a convex function over a convex set is also a global optimal solution (Bazaraa and Shetty (1979, Theorem 3.4.2)). Since the objective function of (4.6) is linear and thus convex, and the set of feasible designs is a convex set by Theorem 4.2., then the Lemma follows.◊

So we are able to say that a local optimal design for the modified SICOED design problem is also a global optimal design (although there may be multiple global optimal designs). This is important because it allows us to stop the optimisation method when we have found a local optimal design. Note that the Lemma holds even when $c(\mathbf{x}_i)\upsilon^2(\mathbf{x}_i)$ is not a convex function of $\mathbf{x}_i$, and also when $\mathcal{X}$ is not a convex set. This is because the $\mathbf{x}_i$ are no longer variables in the problem.

4.9. Non-Linear Programming Solution Method

The simplest approach to numerically solving the SICOED design problem is to apply any one of a large number of well known methods developed for generic non-linear programming problems. This was the approach taken in a number of early papers, e.g. Hartley and Ruud (1969), Box and Draper (1971), Neuhardt and Bradley (1971). The main advantage of this approach is that a nearly "exact" solution to the design problem is obtained, being limited only by the criterion used to stop the optimisation method.

But this approach also has a number of significant disadvantages. First, as seen in the previous section the SICOED design problem does not have the convexity properties required to ensure that non-linear optimisation methods will find the global optimal design. In addition, non-linear solutions methods are likely to behave unpredictably as a result of the presence of asymptotes. Both of these problems can be overcome to some extent by repeating the solution process with different starting values for the design variables.

Second, non-linear programming methods were developed for situations where the cost of evaluating the objective function and constraints are relatively small. As a result they rely on numerous evaluations of the functions that make up the design problem. However, evaluating the design criterion is generally extremely costly, due to a matrix inverse calculation (M^{-1}) and often also an integration calculation (e.g. averaging over the design region). The latter can often be removed from the problem as follows. Many design criteria, such as the average variance of the mean response, can be expressed as $\int_{\mathcal{I}} \mathbf{g}(\mathbf{x})^T \mathbf{M}^{-1} \mathbf{g}(\mathbf{x}) d\mathbf{x}$ where $\mathbf{g}(\mathbf{x})$ is any real vector. To reduce the solution time we can use the fact that (Silvey (1980))

$$\begin{aligned} \int_{\mathcal{I}} \mathbf{g}(\mathbf{x})^T \mathbf{M}^{-1} \mathbf{g}(\mathbf{x}) d\mathbf{x} &= \int_{\mathcal{I}} \mathrm{tr}\left(\mathbf{g}(\mathbf{x})^T \mathbf{M}^{-1} \mathbf{g}(\mathbf{x})\right) d\mathbf{x} \\ &= \int_{\mathcal{I}} \mathrm{tr}\left(\mathbf{M}^{-1} \mathbf{g}(\mathbf{x}) \mathbf{g}(\mathbf{x})^T\right) d\mathbf{x} \\ &= \mathrm{tr}\left(\mathbf{M}^{-1} \int_{\mathcal{I}} \mathbf{g}(\mathbf{x}) \mathbf{g}(\mathbf{x})^T d\mathbf{x} \right). \end{aligned} \tag{4.7}$$

Since $\int_{\mathcal{I}} \mathbf{g}(\mathbf{x}) \mathbf{g}(\mathbf{x})^T d\mathbf{x}$ is a constant, then once it has been calculated at the start of the solution process, (4.7) allows the design criterion to evaluated at each iteration of the solution process without requiring a numerical integration. However there does not appear to be a method for significantly speeding up the calculation of M^{-1}. Hence using non-linear optimisation methods to solve the design problem can require a considerable amount of computer time.

Third, standard non-linear optimisation methods are not designed for optimisation problems with fixed costs or integer variables. For example,

consider the context of steady-state simulation, where Spectral Analysis is used to estimate the variance of the mean response and only one run is performed at each design point. The total number of warm-up observations discarded is then a function of the number of distinct design points, and the length of the initial transient period at those points. However, standard non-linear optimisation methods can generally not deal with such fixed costs of experimentation at each design point.

Lastly, although it would be easier to integrate this approach into an automated design package than an algebraic solution method, it would still not be straightforward to do so. Existing stand-alone packages for non-linear optimisation are not easily integrated into an automated design package, and custom-written code may be lengthy and complex.

One possible solution to the non-convexity problems is to apply the non-linear solution methods to the modified SICOED design problem (4.6). However, this would mean that the number of variables in the problem is increased, being equal to the number factors multiplied by the number of candidate design points. For example, a two-factor problem with an optimal design consisting of 9 design points then has 27 variables for the SICOED design problem, consisting of 9 variables for factor one, 9 variables for factor two, and 9 σ_i^2 variables. For the modified SICOED design problem, with the candidate design points consisting of a grid of 11x11, this example has 121 Σ_i variables. This is a substantial number of variables for any non-linear optimisation method to deal with, and would result in substantially longer solution times.

However this approach does allow very accurate (local) optimal designs to be found. As mentioned before, for design problems with one factor, some or all of the non-convexities in the design criterion can be removed by suitably constraining the design region. For the purpose of further research presented in later chapters, and to allow comparison with the approach shown in the next section, the non-linear programming approach was implemented. Since our primary interest is obtaining optimal designs, and the computer time taken to obtain those designs matters little, the most convenient non-linear programming

method was chosen. Evaluation of the design criterion requires matrix addition, multiplication and inversion. These operations can be very easily performed in the numerical linear algebra package Matlab (*The Mathworks Inc.*), for which a non-linear optimisation routine is also available. Hence Matlab was chosen. Some further details are:

- The non-linear optimisation routine used was the CONSTR routine from the Matlab Optimisation Toolbox. This uses a Sequential Quadratic Programming method (Grace (1990)).
- Simple modifications allow almost any design criterion $L(E^{\sigma})$, any vector function $\mathbf{f}(\mathbf{x})$, any number of factors, and any experimental cost function.
- The variable transformation $\Sigma_i = 1/\sigma_i^2$ was made as this made the solution method faster and more reliable.
- If required, the integration shown in (4.7) is performed using the Matlab routine QUAD, which uses "an adaptive recursive Simpson's rule". For integration in more than one factor, this routine is called recursively.
- For many designs, the matrix inverse required for the design criterion evaluation, M^{-1}, is close to singular. Consideration of such designs can lead to a significantly inaccurate design criterion value. Hence a number of constraint were added to the problem to ensure that the solution routine 'behaved' itself and did not consider solutions that were both clearly non-optimal and likely to lead to a nearly singular M. In particular, the value of $1/\sigma_i^2$ for p design points was constrained to be greater than 0, and the Euclidean distance between the same p design points was constrained to be greater than an arbitrary specified distance.
- The number of design points r was set arbitrarily, usually at $r = p + 1$ or $r = p+2$ unless a larger value of r was suspected. If it is the case that the optimal design consists of less than r points, then we will find that either σ_r^2 is extremely large or there are less than r distinct design points in the design. However the larger r is set to be, the more variables the solution method has to deal with and the more computer time it will take to obtain the optimal design.

Some informal testing was done to assess the effect of the non-convexities in the SICOED design problem on the above solution method. The test case was a two-factor problem, with constant cost-per-experiment and variance functions, using the average variance of the mean response as design criterion, and r = 10. The metamodel was a full quadratic model in both factors. The above method was run 7 times with randomly chosen starting designs, resulting in 3 different designs. The two more costly designs had cost function values that were 1% and 4% above the cheapest design. One of the runs failed to finish because the solution algorithm got 'stuck' at one of the asymptotes in the design criterion function.

4.10. Heuristic Solution Method

Apart from being relatively complex and not guaranteeing global optimality, generic non-linear programming methods were not specifically developed to efficiently solve the design problem and can thus require a significant amount of computer time. The search for more efficient methods has resulted in a substantial literature on 'design algorithms'. These algorithms are heuristics that take advantage of the special characteristics of the design problem to quickly obtain close-to-optimal designs. In this section we will provide a brief review of the literature, and show how the combination and modification of existing design algorithms leads to an efficient design algorithm for the modified SICOED design problem.

Most design algorithms found in the design literature can be seen as variations of the 'greedy' algorithm found in the combinatorial optimisation literature. Robertazzi and Schwartz (1989, p345) explain the basic greedy algorithm, for the combinatorial problem of selecting a subset from a set given an objective function, as follows:

> "The essential characteristic of greedy algorithms is that at each iteration an element that maximises the immediate incremental improvement in the objective function is selected."

In terms of the classical design problem, the 'objective function' is the design criterion, and the 'element' is one experiment added to a particular design point. Since the design criterion is usually a measure of variability, then the addition of one experiment results in a decrease in the criterion value. More specifically, at iteration j a greedy algorithm would add one experiment to the design point $\mathbf{x}_{i*}$ that, over all possible design points, results in the largest decrease in the design criterion value. Letting $\nabla_i L(E_j)$ be the change in design criterion value by adding an experiment at point $\mathbf{x}_i$ to the design E_j at iteration j, we choose

$$\left\{\mathbf{x}_{i*}:\ \nabla_{i*}L(E_j) = \min_{\mathcal{X}} \nabla_i L(E_j)\right\}. \tag{4.8}$$

(note that most design criteria are defined such that $\nabla_i L(E_j) \leq 0$).

To simplify the search for the design point $\mathbf{x}_{i*}$, the design space $\mathcal{X}$ is usually assumed to consist of a finite number of candidate design points, and $\nabla_i L(E_j)$ evaluated at each candidate point. In some situations, the number of candidate points is naturally finite because the factors only have a discrete number of settings. For continuous factors with an infinite number of possible settings, a grid may be used as approximation. Mitchell (1974) lists a number of advantages of restricting the set of candidate points, such as the ability to exclude infeasible or undesirable points, and reducing the complexity of the experiment by keeping the number of possible factor settings low.

Due to the focus of the experimental design literature on D-optimality, most design algorithms have been developed for design problems where D-optimality is the design criterion.

The greedy approach is particularly suitable for the classical design problem where N is small, so that the n_i are integer and a rounded continuous design is not acceptable. By taking a one-experiment step at each iteration, an integer solution is maintained.

In addition to the simplicity of heuristics based on the greedy algorithm, the resulting heuristics may also be considerably faster than non-linear optimisation methods. At each iteration, we need to calculate the value of $\nabla_i L(E_j)$ for each of the potential design points (gridpoints). This results in a large number of design

criterion evaluations. Similar evaluations are required to calculate the derivatives required for the non-linear programming methods. However for the greedy approach we have the following identity. For any matrix A and vector $\mathbf{a}_0$,

$$\left(A+\mathbf{a}_0\mathbf{a}_0^T\right)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{a}_0\mathbf{a}_0^T A^{-1}}{1+\mathbf{a}_0^T A^{-1}\mathbf{a}_0}$$

(Dykstra (1971)). This allows the matrix inversion M^{-1} to be updated from one iteration of a greedy algorithm to another, rather than having to be completely recalculated. If at iteration j+1 we add Δn_i experiments to the n_i experiments at candidate point $\dot{x}_i$, then the above identity implies that

$$M(E_{j+1}) = \left(M(E_j) + \Delta n_i f(\dot{x}_i) f^T(\dot{x}_i)\right)^{-1} = M^{-1}(E_j) - \frac{M^{-1}(E_j)\Delta n_i f(\dot{x}_i) f^T(\dot{x}_i) M^{-1}(E_j)}{1+\Delta n_i f^T(\dot{x}_i) M^{-1}(E_j) f(\dot{x}_i)}.$$

Informal tests show that even for a relatively small problem, updating M^{-1} rather than recalculating it was around 14 times faster.

In terms of the specific implementation of the greedy-type algorithm, there have been two main approaches to solving the classical design problem. One approach (e.g. Wynn (1970), Fedorov (1972)), labelled as *sequential design algorithms*, assumes that the design is (approximately) continuous. Starting from an initial design, experiments are added to design points selected using the greedy algorithm until a stopping criterion is reached, based on the rate of convergence to the optimal design. Once stopped, the design is normalised to give the proportions p_i. This approach is generally considered as an efficient method for determining close-to-optimal continuous designs. However, since experiments are only ever added, there is no facility for removing or reallocating experiments, such as the experiments in the arbitrary initial design. The second approach (e.g. Mitchell (1974), Welch (1982)), labelled as *exchange algorithms*, assumes that an integer design is to be found. This approach starts from an n-point design, and using the greedy algorithm it adds and removes experiments from the design, improving the design criterion value while ensuring that the number of experiments in the design remains at or close to n. Exchange algorithms are

considerably slower than sequential design algorithms, but they produce an integer design that is closer to the optimal design.

Robertazzi and Schwartz (1989) provide a way of speeding up the sequential design algorithm for certain design criteria. The *accelerated sequential design algorithm* assumes that the design criterion has a property called submodularity, which can be described as the combinatorial analogue of convexity. Submodularity implies that the incremental improvement in the design criterion for any candidate point by adding an experiment to that point, $\nabla_i L(E_j)$, is non-decreasing in the number of experiments that make up the design:

$$\nabla_i L(E_j) \leq \nabla_i L(E_k) \quad \forall i, \forall (k > j).$$

Hence if we maintain a record of the most recently calculated value of the incremental improvement in the design criterion value for each point $\dot{x}_i$, which we might label $\nabla_i L(E_{<j})$, then at any iteration j we know that $\nabla_i L(E_{<j}) \leq \nabla_i L(E_j)$.

To illustrate the use of this knowledge, assume that at iteration j we calculate $\nabla_i L(E_j)$ for each point $\dot{x}_i$, and store it in a list. We then add an experiment to the point $\dot{x}_{i*}$ because it satisfies $\min_i \nabla_i L(E_j)$, and determine $\nabla_{i*} L(E_{j+1})$ just for point $\dot{x}_{i*}$. If we find that $\nabla_{i*} L(E_{j+1}) \leq \nabla_i L(E_j) = \nabla_i L(E_{<j})$ for all points $\dot{x}_i \neq \dot{x}_{i*}$, then *without needing to recalculate* $\nabla_i L(E_{j+1})$ *for the remaining candidate points* we know that at iteration j+1 we should again add an experiment to point $\dot{x}_{i*}$. Thus we have performed an iteration with only one design criterion evaluation rather than one evaluation for each candidate point. If we find that another point, say $\dot{x}_{i\wedge}$, was chosen instead of $\dot{x}_{i*}$, then we first update the value in the list corresponding to $\dot{x}_{i\wedge}$, $\nabla_{i\wedge} L(E_{j+1})$, and redetermine whether or not $\dot{x}_{i\wedge}$ was the best candidate point to choose. This process is continued until the best candidate point is chosen, and the next iteration started. Although this algorithm has slightly more overhead than the standard sequential design algorithm, Robertazzi and Schwartz show considerable savings for a number of examples, due to the reduced number of design criterion evaluations required.

It would appear that heuristics based on the greedy algorithm are relatively simple, and may be considerably faster than non-linear programming methods. Hence a suitable heuristic for the SICOED design problem was developed, which is outlined now.

To begin with we assume that the number of candidate design points is finite, say t, and consider the modified SICOED design problem with the variables Σ_i rather than the full SICOED design problem. As seen in section 8 of this chapter, for linear design criteria the modified SICOED design problem has the required convexity properties to ensure that any local optimal solution is also a global optimal solution.

There are two substantial differences between the classical design problem (when the number of candidate design points is t)

$$\begin{aligned} \text{Min} \quad & L(E) \\ \text{s.t.} \quad & \sum_{i=1}^{t} n_i \le N_0 \\ & n_i \ge 0 \quad \text{and integer} \end{aligned} \tag{4.9}$$

and the modified SICOED design problem

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{t} c(\dot{x}_i)\upsilon^2(\dot{x}_i)\Sigma_i \\ \text{s.t.} \quad & L(E^{1/\sigma}) \le L_0 \\ & \Sigma_i \ge 0 \quad \forall i \end{aligned} \tag{4.10}$$

First, the two design problems do not have the same variables. The variables of (4.9) are the number of experiments n_i, which must be integer, while the variables of (4.10) are the inverse of the target variances σ_i^2, which are continuous. Thus the natural step-size for the greedy algorithm, one experiment, is suitable for the classical design problem but not for the modified SICOED design problem. Second, the objective function and (main) constraint of (4.9) are essentially the (main) constraint and objective function respectively of (4.10). However, the cost function of the modified SICOED design problem is a more complex function than the 'sum of n_i' constraint in the classical design problem.

The basic greedy algorithm for the classical design problem consists of adding experiments to minimise L(E) using (4.8). We propose that the basic greedy algorithm for the modified SICOED design problem be as follows:

Iteratively add a stepsize to Σ_{i^*}, associated with design point $\dot{x}_{i^*}$, using the selection criterion

$$\left\{ \dot{x}_{i^*} : \frac{\nabla_{i^*} L(E_j^{1/\sigma})}{c(\dot{x}_{i^*})\upsilon^2(\dot{x}_{i^*})} = \underset{x_i}{\text{Min}} \frac{\nabla_i L(E_j^{1/\sigma})}{c(\dot{x}_i)\upsilon^2(\dot{x}_i)} \right\} . \tag{4.11}$$

Thus an experiment is added to the candidate point that gives the best change in the design criterion *per unit cost*. This is sufficient to describe the equivalent of the classical sequential design heuristic described above. For the classical exchange algorithms, experiments are both added and removed, requiring the rule that the sum of the n_i must be at or close to N_0. For the modified SICOED design problem, we change this to the rule that *the design criterion value must be at or near the target L_0.*

For any greedy-type design algorithm, the amount of computer time taken to determine an 'optimal' design depends mainly on the number of candidate design points. Thus it would seem sensible to split the process of finding an optimal design into two stages: (i) finding the subset of the candidate design points for which we believe that $\Sigma_i > 0$ in the optimal design, and (ii) finding the optimal values of the Σ_i in this subset.

Rather than using either the sequential design algorithm or the exchange algorithm, we suggest a combination of modified versions in three phases, as shown in Figure 4.5. The main idea behind the 3-phase algorithm is to use the respective strengths of the sequential and exchange algorithms to achieve the results desired at various stages of the design algorithm.

The objective of Phase 1 is to quickly determine the candidate design points that are likely to appear in the optimal design, and find a rough estimate of the associated values of Σ_i. A modified sequential design algorithm is used. In Phase

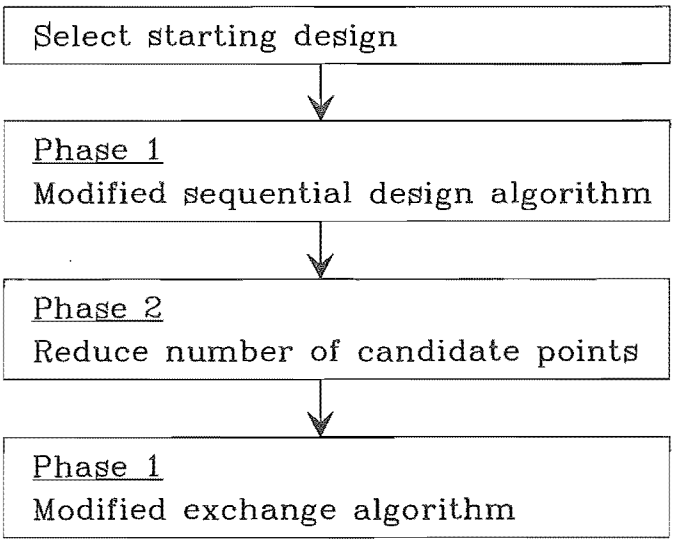


Figure 4.5. Three-phase design algorithm for modified SICOED design problem

2, we use various heuristics to remove those candidate design points that are unlikely to appear in the optimal design. This means that in Phase 3, a modified version of the exchange algorithm has to deal with a much smaller number of candidate design points.

The specific heuristics used for each phase are as follows:

Starting Design: Either a random starting design may be used, or a specific starting design selected by the experimenter, such as a factorial design.

Phase 1 - Sequential design algorithm as described above, with the following modifications:

- The step-size for any candidate design point $\dot{x}_i$ at any iteration is set at max(1, $0.1\Sigma_i$). The number 1 is arbitrary, as there is no 'natural' step-size for the modified SICOED design problem. In some cases, the cost-per-experiment function is such that a large number of experiments should be performed in one part of the design region, and only a few in the rest of the design region. To ensure that the algorithm does not spend most of the iterations adding the

step-size to only one or a few candidate points, the stepsize is set at the greater of 1, and 10% of the current value of Σ_i.

- The candidate point to add the step-size to, is chosen using (4.11). $\nabla_i L(E_j^{1/\sigma})$ is calculated using the step-size 1.
- To speed up the algorithm, the accelerated version of the sequential design algorithm can be used, provided the design criterion used has the submodularity property described before.
- The stopping rule for Phase 1 should be related to the objective of this phase: To find the candidate points that appear in the optimal design. The minimum number of points in the design is p, so a simple heuristic rule is to stop when the there are p points that have been added to 4 times, i.e. the value of the p^{th} largest Σ_i is 4. This heuristic ensures that in most cases, any points that are likely to be in the final design have $\Sigma_i > 0$ at the end of Phase 1.

Phase 2 - reducing the number of candidate design points using the following heuristics:

- Each candidate point for which $\Sigma_i = 0$, is removed.
- If $\Sigma_i < \Sigma_j$ and candidate point i is an immediate neighbour of candidate point j in Σ_i space (assuming some sort of grid is used to define the candidate design points), then candidate point i is removed, provided neither i nor j have previously been removed. All combinations of candidate design points are examined sequentially in this way. Often the optimal design point at which to perform experiments lies between two candidate design points. The result is that both candidate points are added to in Phase 1. This rule ensures that only one of the two points is considered in Phase 3.

At the end of Phase 2 the design is usually reasonably efficient, but so far no notice has been taken of the actual magnitude of the design criterion value. Before Phase 3 begins we need to ensure that the design is transformed so that the design criterion value is (approximately) equal to L_0. If the design criterion is a linear criterion (as defined in section 6), then by definition $L(cE^{1/\sigma}) = cL(E^{1/\sigma})$, and hence the Σ_i values can simply be scaled by the ratio $L_0 / L(E_{phase2}^{1/\sigma})$. This scaling does not impact on the efficiency of the design in terms of the design

criterion, since only the relative values of Σ_i matter. However if the cost function is not linear, for example if the cost function is a step function, then efficiency may be affected.

Phase 3 - Exchange algorithm as described above, with the following modifications:

- Instead of ensuring that the design consists of a certain number of experiments, the algorithm must ensure that the design criterion value is close to L_0. Hence when $L(E_j^{1/\sigma}) > L_0$, the step size at iteration j+1 is positive, and when $L(E_j^{1/\sigma}) < L_0$ the step size at iteration j+1 is negative.
- As in Phase 1, the step-size must be set arbitrarily. However, the step-size must also be reduced as the algorithm progresses, to allow a more accurate design. Thus the step-size is reduced slightly at every iteration. A simple multiplying constant is used, such as 0.95 or 0.99.
- The step-size is allowed to be different for each candidate point, reflecting the different magnitudes of the Σ_i values.
- To prevent cycling, we significantly reduce the step-size for a particular candidate point when the algorithm has first added, and then subtracted from that point in two successive iterations.
- To speed up the algorithm, we significantly increase the step-size for a particular candidate point when the algorithm has, in two successive iterations, added (subtracted) to (from) that point.
- The candidate point to add (subtract) the step-size to (from), is chosen using (4.11). The value of $\nabla_i L(E_j^{1/\sigma})$ is calculated using the step-size $1^{-5}*\max(\Sigma_i)$, appropriately signed according to whether $L(E_j^{1/\sigma}) > L_0$ or $L(E_j^{1/\sigma}) < L_0$.
- The stopping rule can be any convergence rule, such as convergence of the cost function value, or a minimum step size.

A number of other heuristic rules that could be used as part of these phases were also developed, but these either increased the time taken to converge, or resulted in poor selection of the candidate points for Phase 3:

- A design algorithm consisting solely of Phase 3 was considered, as this allows more accurate designs to be found. However this proved to be very slow.

- To calculate the value of $\nabla_i L(E_j^{1/\sigma})$, we could use the actual step-size to be taken at that iteration (which is generally different for each candidate point) rather than a small common increment (±1 for Phase 1, ±0.0001*max(Σ_i) for Phase 3). However, this rule actually increased the time taken to converge.
- Dynamic step-size: The use of a discrete heuristic for a continuous problem would appear to be contradictory. However, the main focus of the heuristic is not to optimise, but to goal-seek. In Phase 3, the heuristic behaves in a saw-tooth fashion, by jumping over and under the criterion target L_0 using discrete steps, improving the design efficiency at each step and generally taking smaller steps as the iterations progress. This explains why dynamically calculating the step-size at each iteration makes little sense.
- Note that the method used to accelerate the sequential design algorithm cannot be applied throughout the exchange algorithm of Phase 3. This is because both positive and negative steps are taken, and the list of $\nabla_i L(E_{<j}^{1/\sigma})$ values is changed almost completely by the change in sign of the step. It can be applied when two consecutive positive or negative steps are taken, but the extra overhead does not appear to make this worthwhile.

An example of the design points selected by each phase of the 3-Phase heuristic for a particular design problem is shown in Figures 4.6., 4.7. and 4.8. The design problem has 2 factors, constant cost-per-experiment and variance functions, and the metamodel was a full quadratic metamodel in both factors. The design criterion was the average variance of the mean response over the design region. An 11x11 grid was used over the square design region $\mathcal{X} = \{\mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\}$.

Phase 1, which took 11.37 seconds on a 486dx2-66, shows that relatively few of the 121 gridpoints are likely to be in the final design. Only 22 candidate points have $\Sigma_i > 0$. Of these, 13 remain as candidate points after Phase 2, which took 0.55 seconds. The output of Phase 3, which took 9.34 seconds, shows that the 9 design points in the final design are positioned like a factorial design.

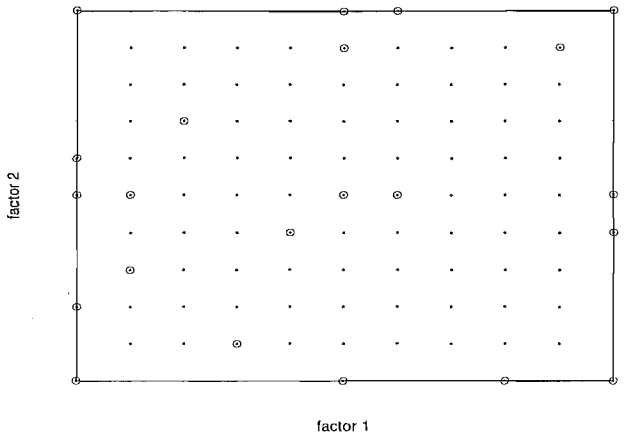


Figure 4.6. Output of Phase 1 ('·' indicates gridpoint, 'o' indicates $\Sigma_i > \vec{0}$)

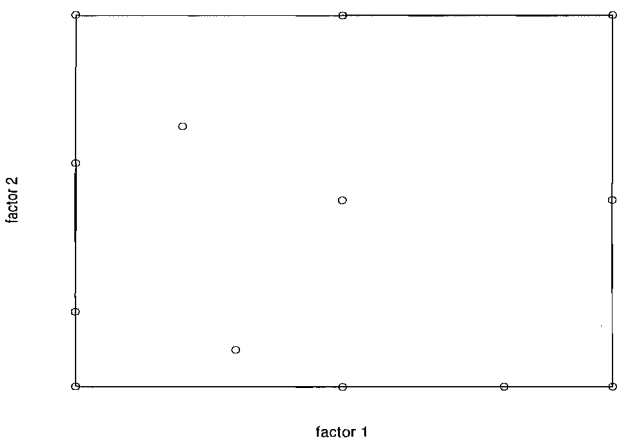


Figure 4.7. Output of Phase 2 ('o' indicates a candidate design point)

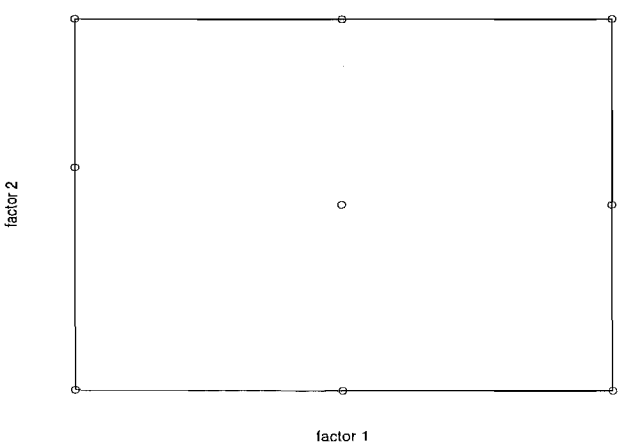


Figure 4.8. Output of Phase 3 ('o' indicates design point)

The main advantage of the 3-phase heuristic is that it is relatively simple to code-up in software, and incorporate into or attach to other software packages. Experience shows that the solution time taken by the heuristic for a close-to-optimal solution is significantly smaller than generic non-linear optimisation methods (often by an order of magnitude), and depends largely on the accuracy required. A feasible design is very quickly found, but once a close-to-optimal design is found it can take a large number of iterations to substantially improve on it.

However, when there are a very large number of candidate design points, the heuristic may take an unreasonably long time to complete. For example, a situation with 4 factors and a grid with 11 settings for each factor leads to 14641 candidate design points. In such situations, sections of the design region could be removed, or a coarser grid used.

Note that the design found using the heuristic is not optimal for the original SICOED design problem of the experimental situation, because the number of candidate design points is assumed to be limited.

4.11. Summary

In this chapter we have considered the SICOED design problem from an implementation perspective. The options available for choosing specific components of the design problem were examined, and some detailed suggestions made for the design criterion and cost function, and parameter estimation method.

Once the design problem is completely specified, there are three main approaches for finding the optimal experimental design. The first is an algebraic approach. However, this approach is only suitable for use by experimenters with a high level of mathematical knowledge, and it would not appear possible to integrate this approach into experimental design software.

The second approach is to apply standard non-linear programming methods. Unfortunately, the SICOED design problem does in general not have the

convexity properties required to ensure that a local optimal design is also a global optimal design. Thus non-linear programming methods cannot guarantee global optimality. In addition those methods are relatively slow, as they require many evaluations of the design criterion, and may fail to converge due to the shape of the design criterion function.

The third approach is to apply a heuristic solution method. Rather than finding a solution to the SICOED design problem, we apply heuristic methods to a modified SICOED design problem, where the number of candidate design points is finite. The advantage of the modified design problem is that it has the required convexity properties to ensure that a local optimal design is also a global optimal design. The 3-Phase heuristic design algorithm developed in section 10 is able to quickly find a close-to-optimal design. This heuristic is a combination of modified versions of the sequential and exchange algorithms found in the classical design algorithm literature, together with a number of additional heuristic rules.

CHAPTER 5: PROPERTIES OF THE SICOED APPROACH

- Examples and a Monte Carlo Study -

5.1. The Distribution of Information Across $\mathcal{X}$

In this section we present an example of the effect that misspecifying the variance function has on the distribution of information across the design region, for both a classical factorial design and a design found using our "Semi-sequential Information Constrained Optimal Experimental Design" (SICOED) approach. The example consists of an experimental situation with one factor (x), and one response (y) which is normally distributed. The variance function of the response is $\upsilon^2(x) = 2x + 1$. To focus on the effect of the variance function we will assume that the cost-per-experiment function is constant, and so we set $c(x) = 1$. The design region is $\mathcal{X} = \{0,1\}$, and the second order polynomial metamodel

$$\begin{aligned} y(x) &= \mathbf{f}^T(x)\beta \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 \end{aligned}$$

is to be fitted over this region. Hence $\mathbf{f}(x) = \begin{bmatrix} 1 & x & x^2 \end{bmatrix}^T$. The design criterion to be used is the average variance of the fitted response over $\mathcal{X}$ (see section 5 of Chapter 4), with a target value of $L_0 = 0.01$. The Estimated Weighted Least Squares estimators will be used, and since the designs found below have the property that $r = p$ then these estimators are unbiased (see section 4 of Chapter 4).

The design problem for this situation can then be stated as

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{r} \frac{2x_i + 1}{\sigma_i^2} \\ \text{s.t.} \quad & \int_0^1 f^T(x)\left(\sum_{i=1}^{r} \frac{1}{\sigma_i^2} f(x_i) f^T(x_i)\right)^{-1} f(x)dx \Big/ \int_0^1 dx \le 0.01 \qquad (5.1) \\ & 0 \le x_i \le 1 \quad \forall i \\ & \sigma_i^2 \ge 0 \quad \forall i \end{aligned}$$

To determine the optimal design, we apply the non-linear programming method described in section 9 of Chapter 4. As suggested there, we add the constraint $x_1 < x_2 < ... < x_r$ to prevent a singular M matrix. By using several starting designs, the optimal design

$$x_1 = 0, \qquad x_2 = 0.4702, \qquad x_3 = 1,$$
$$\sigma_1^2 = 0.01282, \qquad \sigma_2^2 = 0.01, \qquad \sigma_3^2 = 0.02157$$

is obtained (this design was checked against the design found using the 3-phase heuristic method proposed in section 10 of Chapter 4, to ensure that it was (close to) globally optimal). Hence the experimenter would, for example, perform enough experiments at design point x_1 to ensure that the estimated variance of the mean response at x_1 was no more than 0.01282. Note that compared to a 3-level factorial design, the above design is slightly different in that the middle design point lies at x = 0.47 rather than x = 0.5. Also, the above design concludes that a significantly more accurate estimate of the mean response should be collected at x = 0, where the variance of the responses is smallest, than at x = 1, where the variance of the responses is largest.

For the purpose of this example, we assume that (i) any estimators used are unbiased, (ii) that there is no integer restriction on the n_i, and (iii) that the sequential sampling procedure employed is such that the expected value of the estimated variance of the mean response, s_i^2, is equal to σ_i^2. Hence the expected number of experiments at each point x_i is given by $E[n_i] = \upsilon^2(x_i) / \sigma_i^2 = (2x_i+1) / \sigma_i^2$, and we can determine that

$$E[n_1] = 77.98, \qquad E[n_2] = 194.04, \qquad E[n_3] = 139.09.$$

The expected total number of experiments to be performed, for the SICOED approach and using the correct variance function, is thus 411.10. Since the n_i are sufficiently large, the relaxation from being integer will have little impact on the conclusions drawn from this example.

To illustrate a point made in Chapters 2 and 3 regarding the effect of misspecifying the variance function, let us now assume that the experimenter has mistakenly assumed that the variance of the responses was constant, rather than linearly increasing in x. Hence we set $\hat{\upsilon}^2(x) = 1$ in the design problem (5.1). The solution to (5.1), using the same solution method as before, is now

$$x_1 = 0, \qquad x_2 = 0.5, \qquad x_3 = 1,$$
$$\sigma_1^2 = 0.0175, \qquad \sigma_2^2 = 0.01, \qquad \sigma_3^2 = 0.0175.$$

Note that this is closer to a factorial design, although this design requires a more accurate estimate of the mean response at the middle design point than a factorial design does.

Using the true variance function, $E[n_1] = 57.14$, $E[n_2] = 200$, $E[n_3] = 171.43$, and $E[n_1 + n_2 + n_3] = 428.57$. As expected this total is slightly higher than for the SICOED design using the correct variance function. Looking at the design shown above, we see that $\sigma_1^2 = \sigma_3^2$ and that x_2 lies midway between x_1 and x_3. Hence we would *expect* that the variances of the fitted response obtained from the experiments would be symmetric around x_2. This is shown in Figure 5.1., where the solid line shows the expected variance of the fitted response across the design region, resulting from the above design.

The classical optimal design, corresponding to the above situation where the variance function is (wrongly) assumed to be constant, is

$$p_1 = {}^4/_{15}, \qquad p_2 = {}^7/_{15}, \qquad p_3 = {}^4/_{15}.$$

This can found by noting that $p_i \propto \hat{\upsilon}^2(x) / \sigma_i^2$, and $\Sigma p_i = 1$. Let us assume that the classical optimal design / sequential analysis combination (see section 4 of Chapter 3) is used, and enough repetitions of the above classical optimal design are performed in a sequential manner in order to meet the design criterion. Using the true variance function we have $\sigma_i^2 = (2x_i+1) / (p_i E[N])$, and by substituting this into L(E) we can determine that we must have $E[N] = 428.57$ to have L(E) equal 0.01 as required. Hence sequential repetition of the classical design results in

$$E[\sigma_{c1}^2] = 0.00875, \quad E[\sigma_{c2}^2] = 0.01, \quad E[\sigma_{c3}^2] = 0.02625.$$

For this example, with a mis-specified variance function, the actual experimental cost for the SICOED approach is the same as for the optimal design / sequential analysis approach, 428.57. However, although the design points are identical, the amount of information collected at each design point is substantially different, as shown by the σ_i^2 values. The expected variance of the fitted response for the optimal design / sequential analysis combination is also shown in Figure 5.1., as a dashed line.

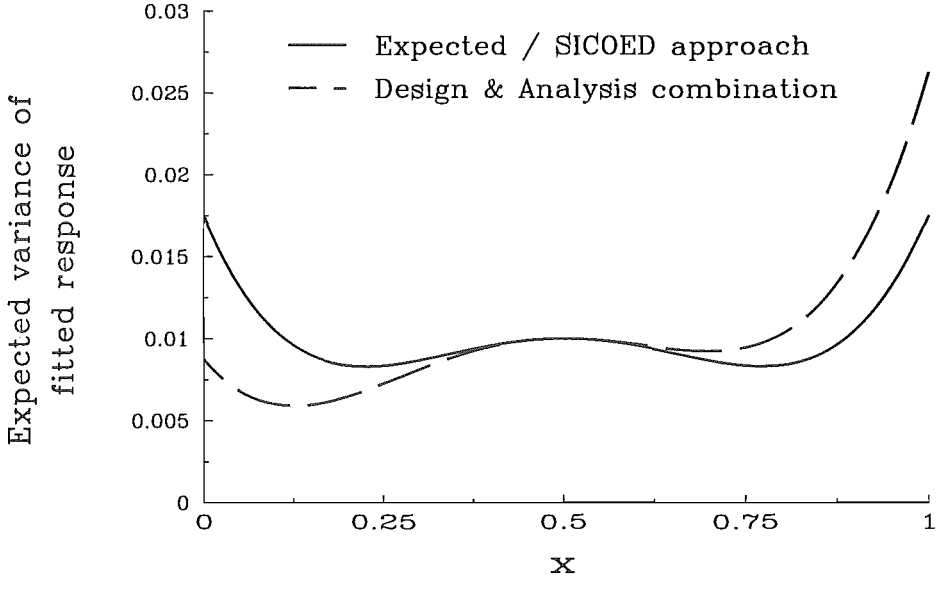


Figure 5.1. Comparing the information spread for two approaches

Figure 5.1. illustrates a point made in Chapters 2 and 3; that the combination of optimal design and sequential analysis may lead to a spread of information that is substantially different from that expected. In fact for a classical optimal design, the value of any arbitrary design criterion (including the one used to determine the design) will in general be higher or lower than the value of the same criterion based on the actual data collected, unless the variance function is known exactly. Generally, the worse the estimate of the variance function is, the bigger this difference becomes. On the other hand, the SICOED approach ensures that the

expected spread of information is obtained. This is because the number of experiments performed at each design point is not fixed. Instead, the design consists of stopping rules.

5.2. Comparing the Cost of Various Design Methods

The SICOED approach has a number of features that make it more flexible, and more suitable for use in experimental design software, than classical approaches. However the SICOED approach can also be significantly more efficient. In this section we consider a simple example - steady-state discrete-event simulation of the M|M|1 queuing system - to highlight the differences in experimental cost between the SICOED approach and various other approaches found in the literature. We will consider the experimental design problem for a steady-state simulation model of the M|M|1 queue, with one factor, the traffic intensity (ρ), and one response, the mean queue length (L_q). The design region is the traffic intensity between $\rho_L = 0.6$ and $\rho_U = 0.8$, over which we wish to fit the second order polynomial

$$L_q = \beta_1 + \beta_2\rho + \beta_3\rho^2$$

(such a model can provide a very good representation of the true mean response over this region, ensuring that specification error is negligible). Hence in terms of the metamodel we have $\mathbf{f}(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 \end{bmatrix}^T$ and $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^T$. The design criterion to be used is the average variance of the fitted response (see section 5 of Chapter 4), using Estimated Weighted Least Squares (note that the designs found below have r = p, so the EWLS estimators are unbiased, see section 4 of Chapter 4), and with $L_0 = 0.02$. Finally, the cost per experiment $c(\rho)$ is approximately constant in discrete event simulation of this queue (Cheng and Kleijnen (1995)), and we assume that the same length of warm-up period is used for all runs.

Seven different approaches to the problem of finding an appropriate experimental design for this situation will be considered, labelled OPTIMAL, LINEAR, LINEAR (1/2), LINEAR (2), CLASOPT, FACTORIAL, and CONSTVAR. Four of these consist of our approach with different cost functions, and the remaining

three are existing methods discussed in Chapter 2. Details of the approaches are as follows:

Optimal design, SICOED approach *(OPTIMAL, LINEAR, LINEAR (1/2), LINEAR(2))*

Four cases are presented, corresponding to three levels of the experimenter's knowledge. The first, *OPTIMAL*, assumes that the experimenter knows what the variance function is (up to a constant of proportionality). For our example, the actual asymptotic variance function is

$$\upsilon^2(\rho) = 2\rho^2\left(1+4\rho-4\rho^2+\rho^3\right)/(1-\rho)^4,$$

(Whitt (1989)), leading to the design problem

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{r} \frac{2\rho_i^2\left(1+4\rho_i-4\rho_i^2+\rho_i^2\right)}{\sigma_i^2(1-\rho_i)^4} \\ \text{s.t.} \quad & \int_{0.6}^{0.8} f^T(\rho)\left(\sum_{i=1}^{r}\frac{1}{\sigma_i^2}f(\rho_i)f^T(\rho_i)\right)^{-1} f(\rho)d\rho \Big/ \int_{0.6}^{0.8} d\rho \le 0.02 \\ & 0.6 \le \rho_i \le 0.8 \quad \forall i \\ & \sigma_i^2 \ge 0 \quad \forall i \end{aligned} \tag{5.2}$$

The second, *LINEAR*, assumes that the experimenter knows the exact value of the variance function at the two ends of the design region, $\upsilon^2(\rho_L)$ and $\upsilon^2(\rho_U)$, and uses a linear function to interpolate between these points (see Figure 5.2.).

Two further cases, *LINEAR (1/2)* and *LINEAR (2)*, assume that the experimenter makes a significant error in estimating the variance function by linear interpolation. Specifically, the ratio of $\upsilon^2(\rho_L)$ to $\upsilon^2(\rho_U)$ is taken to be half (*LINEAR (1/2)*) or double (*LINEAR (2)*) the correct ratio. The design corresponding to each of these cases can be found by solving the design problem (5.2), using the same approach as for the example in section 1, with the appropriate variance function. Note that in all cases, the optimal number of design points r was 3.

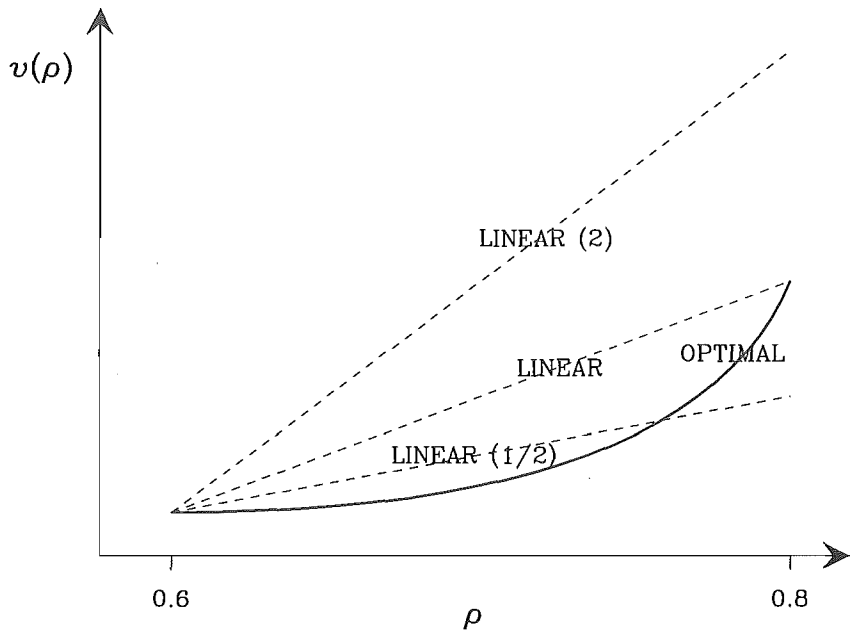


Figure 5.2. The variance function approximations used for the SICOED designs

Optimal design, classical approach *(CLASOPT)*

For this classical optimal design approach, $\upsilon(\rho)$ is taken to be a constant function, reflecting the classical assumption of a common error variance. The design is found as above.

Constant run-length factorial design *(FACTORIAL)*

This is the type of design seen most often in the simulation literature. It consists of design points at $\rho = 0.6$, 0.7 and 0.8, with an equal number of observations n_i collected at each of these points. To obtain the design in terms of the required variances, we use the relationship $E[n_i] = \upsilon^2(\rho_i) / \sigma_i^2$ (which assumes n_i is continuous). Since $n_1 = n_2 = n_3$, then we must set $\sigma_i^2 \propto \upsilon^2(\rho_i)$ with the same constant of proportionality for each i. To ensure that the design can be compared to the 'optimal' designs, the correct value of this constant is the one for which the design criterion target L_0 is exactly satisfied.

Constant variance factorial design *(CONSTVAR)*

This design is similar to the *FACTORIAL* design, except that the run-lengths are adjusted so that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ (see section 3 of Chapter 2). This implicitly adds a second design criterion. As for the *FACTORIAL* design, the value of σ_i^2 is determined by ensuring that the design criterion target is exactly satisfied.

Note that for comparison purposes, two additional assumptions have been made regarding the designs based on the classical framework (*CLASOPT*, *FACTORIAL* and *CONSTVAR*). First, we assume for all the designs that a single long run is performed at each design point, so that the variance function is known and no additional complexity is added due to the varying number of warm-up periods used by each design. Clearly this is incompatible with classical designs, which in practice require a fixed number (greater than 1) of replications of pre-determined length. However, this assumption leads to an underestimate of the expected experimental cost for these designs, as it is more efficient in the case of the M|M|1 queue to perform one long run than multiple independent replications (Whitt (1991)). Second, for these classical designs it is assumed that the design has been found by setting the run-lengths so that the design criterion is exactly satisfied, allowing comparison with the optimal designs. The experimental design literature does not guide the experimenter in the choice of an appropriate run-length, and unless such a procedure is used, the actual outcome of the experiment is likely to be significantly different from that desired.

In addition to these 'best-case' scenario assumptions for the *CLASOPT*, *FACTORIAL* and *CONSTVAR* designs, the comparison does not take into account the effect of the variance estimation technique used, which for these designs is limited to the potentially inefficient method of independent replications.

The resulting designs are shown in Table 5.1. The last column of that table contains the true expected cost of each design (relative to that of the *OPTIMAL* design), found by evaluating each design using the cost function associated with the *OPTIMAL* design problem.

Approach	*Design points*	*Required Variances*	*Expected Cost**
OPTIMAL	0.6 0.677 0.793	0.0154 0.0124 0.0614	1
LINEAR	0.6 0.682 0.8	0.0113 0.0184 0.0553	1.108
LINEAR (1/2)	0.6 0.684 0.8	0.0153 0.0184 0.0534	1.109
LINEAR (2)	0.6 0.681 0.8	0.01 0.0183 0.0560	1.112
CLASOPT	0.6 0.7 0.8	0.0375 0.0188 0.0375	1.441
FACTORIAL	0.6 0.7 0.8	0.0032 0.0140 0.0910	1.328
CONSTVAR	0.6 0.7 0.8	0.025 0.025 0.025	1.915

* Found by evaluating each design using the cost function associated with the *OPTIMAL* design problem, and dividing this by the cost for the *OPTIMAL* design

Table 5.1. Comparing various design methods

A number of conclusions can be drawn from the differences between the designs shown in Table 5.1. Interestingly, the designs for *LINEAR*, *LINEAR (1/2)* and *LINEAR (2)* are very similar in all respects. However, the cost of the *CLASOPT* design is considerably higher than for these designs, indicating that the important thing is to have a variance function with positive slope. The actual size of the

slope of the variance function appears to make little difference for a reasonably wide range of values.

For this example, it is clear that the SICOED approach performs significantly better than the existing design methods. This is achieved through the use of a design problem, and by incorporating information contained in the (estimated) variance function, into this design problem. Even very rough estimates of the variance function, such as in the *LINEAR (1/2)* and *LINEAR (2)* cases, lead to only a relatively small penalty over the *OPTIMAL* design. It is also interesting to note that the factorial design outperforms the classical optimal design, indicating that an 'optimal' design based on very inaccurate information is likely to perform badly. On the other hand, the implicit design criterion of constant variance used to obtain the CONSTVAR design clearly results in a large cost penalty over the other designs.

5.3. Jackson Queueing Network with Unknown Marginal Cost Function

The two examples presented in sections 1 and 2 had known variance functions, so that the designs could be evaluated without performing any actual experiments. This allowed a number of points to be illustrated without the need to run numerous simulations. For the example in this section, simulation of a Jackson queueing network with overtaking, there is no theoretical result for the form of the variance function. This example illustrates the steps that are usually required to determine a design for an arbitrary experimental situation. We also run the simulations required by each design and analyse the results. A number of observations can be made about the SICOED approach from the simulation data produced.

All simulations and computations reported below were performed on an Intel 486DX2-66.

Figure 5.3. shows a simple 3-node Jackson network, which has been the subject of some interest in the queueing literature (e.g. see Lemoine (1979), Simon and Foley (1979)), and whose properties have also been investigated using

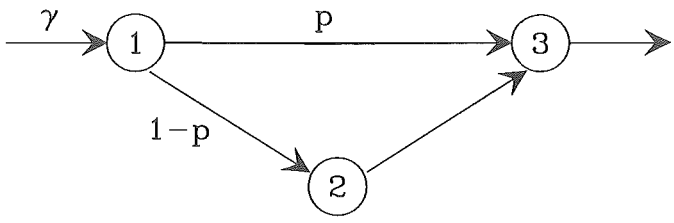


Figure 5.3. Jackson network example

simulation (Kiessler and Disney (1982)). Customers arrive at node 1 according to a Poison arrival process with rate γ. After service, customers proceed to node 3 with probability p or node 2 with probability 1-p. Those customers who travel through node 2 then proceed directly to node 3. At each node there is a single server, and service times are negative exponentially distributed with a mean service time of 1. We wish to determine the relationship between the mean time in the system, W, and the factors γ and p.

From queueing theory we know that the mean number of customers in the system is given by

$$\begin{aligned} L &= L_1 + L_2 + L_3 \\ &= \frac{\gamma}{1-\gamma} + \frac{\gamma(1-p)}{1-\gamma(1-p)} + \frac{\gamma}{1-\gamma} \\ &= \frac{2\gamma}{1-\gamma} + \frac{\gamma(1-p)}{1-\gamma(1-p)} \end{aligned}$$

and hence the mean time in the system (from Little's formula) is

$$\begin{aligned} W &= \frac{L}{\gamma} \\ &= \frac{2}{1-\gamma} + \frac{1-p}{1-\gamma(1-p)}. \end{aligned} \tag{5.3}$$

Although the mean time in the system is known, the steady-state distribution (and hence the variance) of W has not yet been derived (see Simon and Foley (1979)).

However, without knowledge of these theoretical results we could simulate this system, and fit a metamodel to the responses. Cheng and Kleijnen (1995) investigate a similar situation, and suggest that the general metamodel

$$y(x) = (\beta_0 + \beta_1 x + \ldots + \beta_k x^k) g(x) + \varepsilon$$

can be used for queueing systems with queue saturation. The saturation effect can be modelled by appropriately selecting the function g(x). For our network, we assume that a first-order model in γ and p is adequate for the polynomial component in the metamodel. Since queue saturation occurs as $\gamma \rightarrow 1$, we set g(x) = $1/(1-\gamma)$, leading to the metamodel

$$W(\gamma, p) = (\beta_0 + \beta_1 \gamma + \beta_2 p) \frac{1}{1-\gamma}.$$

In fact, this metamodel is able to provide a very good representation of (5.3) over the design region selected below, ensuring that any bias introduced by incorrect metamodel selection is negligible.

We will assume that the objective of the experiment is to determine a metamodel for W over the design region $\mathcal{X} = \{\gamma, p: 0.8 \leq \gamma \leq 0.95, 0.25 \leq p \leq 0.75\}$, which includes the most frequently studied 'load' range (0.8 to 0.95) of such a queueing system. Two approaches to the determination of an experimental design for the simulation runs will now be considered.

Classical approach

Since there are three parameters (β_0, β_1 and β_2) and two factors (γ and p) in the metamodel, we can use a 2^2 factorial design. The factorial design is then made up of the design points (0.8, 0.25), (0.8, 0.75), (0.95, 0.25), and (0.95, 0.75), as shown in Figure 5.4.

As is standard in the simulation design literature, we will assume that an equal number of runs are performed at each design point, and that the run-length is the same for each run. In this case (an arbitrary) 20 runs are performed at each

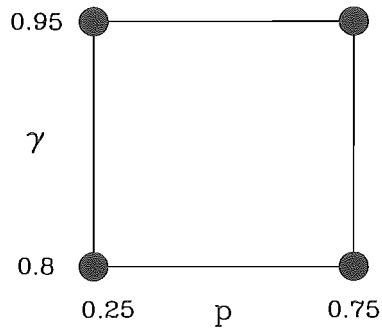


Figure 5.4. Classical factorial design

point, so that an estimate of the pure error at each point can be made. From each run we collect only the mean time in the system. The run-length is set to be 20,000 customers, preceded by a 'warm-up' run-length of 10,000 customers. This experiment (20 runs at 4 design points) was repeated 30 times with different random number offsets. Some relevant data is shown in Table 5.2.

γ	p	*Runs per Replication*	*Total Cost (sec)*	W *(from (5.3))*	$\overline{W}(\gamma,p)$ *(simul.)*	*Mean* $\hat{V}ar(\overline{W}(\gamma,p))$
0.8	0.25	20	12421	11.875	11.88	0.01484
0.8	0.75	20	10894	10.313	10.31	0.01354
0.95	0.25	20	19900	42.609	42.50	3.568
0.95	0.75	20	17771	40.328	40.67	4.037

Table 5.2. Data from 30 repetitions of the classical design

Comparing the values of W from (5.3), and the $\overline{W}(\gamma,p)$ values from simulation, it would appear that the simulation model is a valid model of the Jackson network in Figure 5.3. Note that the cost of the experiment is measured as seconds of computer time, and includes the cost of the warm-up period. The cost shown is for all 30 experiments. $\hat{V}ar(\overline{W}(\gamma,p))$ is the variance of the mean time in the system calculated from the 20 runs performed at (γ, p). The last column in the table shows the *mean* of the 30 variances determined at each design point from

the 30 replications. The actual values making up this mean can be found in Table A1.1. of Appendix 1.

From Table 5.2. it can be seen that there are large differences in the variance of the mean response across the design points, mainly due to the queue saturation effect at high values of γ. Hence when fitting the metamodel we should not assume constant variance, as is usually assumed for the response data from a factorial design. Also, there is a significant difference in the computer time required at various levels of γ and p. The effect of changing p on the cost of the experiment is easily explained, since a smaller p will result in a larger number of customers travelling though 3 instead of 2 nodes, resulting in a larger average number of events per customer. The effect of γ on the cost is likely to be due to the extra overhead at higher loads, when queue lengths are longer.

These observations suggest that an optimal experimental design may be significantly more efficient than the factorial design used above.

Assume now that the experimenter's objective is quite general - essentially obtaining a metamodel that provides a reasonable estimate of the response across the design region. Estimated Weighted Least Squares will be used to determine the parameters of the metamodel, and since the number of design repetitions is 20 then the bias in the parameter variance estimates is likely to be reasonably small (see section 4 of Chapter 4).

A suitable design criterion for this objective is the average variance of the fitted response over the design region. For the factorial design used above, the average value of this criterion over the 30 experiments is 0.265, with a standard deviation of 0.09. The actual values can be found in Table A1.1. of Appendix 1. That table shows that there is quite a range of design criterion values that results from a fixed number of experiments, in this case from 0.1485 to 0.5114. It appears that even an average over 20 runs of 20,000 customers each is still quite variable.

SICOED Approach

In order to allow comparison between the classical approach and the SICOED approach, we will use the same design region and design criterion. The design criterion target is set at 0.265, being the mean of the design criterion values resulting from the factorial design. The same method of performing simulations is also used, consisting of the method of independent replications with run-lengths of 20,000 customers after a 10,000 customer warm-up period. As for the factorial design, we will also perform 30 replications of the SICOED approach, although this time there will be 30 different designs.

For the SICOED approach we first require estimates of the cost-per-experiment function $c(\gamma,p)$ and the variance function $\upsilon^2(\gamma,p)$ across the design region. As shown in section 3 of Chapter 4, it is sufficient to obtain an estimate of the function $c(\gamma,p)\upsilon^2(\gamma,p)$. We will refer to this as the marginal cost function. This marginal cost function estimate, for various (γ,p), may in general be obtained from the verification and validation stages of constructing the simulation model, previous experiments with this simulation model, or a pilot experiment.

In this case, we will assume that no data is available, so that a pilot experiment is performed. Two strategies can be employed with respect to the data obtained from a pilot experiment: (1) We get a quick estimate, and discard the pilot data (e.g. because the warm-up period was too short to be sure that steady state had been reached), or (2) we aim to use the pilot data in the final analysis. For this example, due to large warm-up period required at higher loads, the former strategy will be used. The pilot was chosen to consist of 3 runs of 10,000 customers each (preceded by a warm-up run of 5,000 customers) at the points of a 2^2 factorial design plus a centre point as shown in Figure 5.5. Note that the ratio of warm-up to run-length is as required by the comments made in section 3 of Chapter 4, so that a reasonable estimate of the marginal cost function is obtained. This pilot experiment was then performed 30 times (with different random number seeds), to obtain 30 marginal cost functions. A summary for the 30 pilot experiments is shown in Table 5.3.

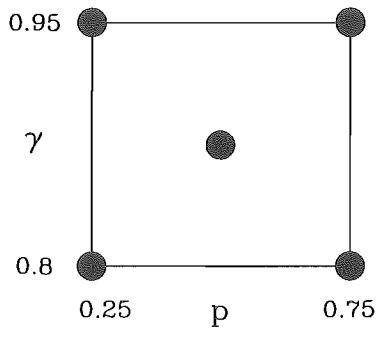


Figure 5.5. Design for pilot experiments

γ	p	*Mean cost (seconds)*	*Mean* $\hat{\mathrm{Var}}(\overline{W}(\gamma,p))$	*Mean* $c(\gamma,p)\upsilon^2(\gamma,p)$
0.8	0.25	30.9	0.148	4.57
0.8	0.75	27.2	0.185	5.03
0.875	0.5	32.6	1.27	41.2
0.95	0.25	49.5	39.3	1948
0.95	0.75	43.9	57.4	2521

Table 5.3. Data from pilot experiments

A complete listing of the estimated marginal cost function values at the 5 design points for the 30 experiments can be found in Table A1.2. of Appendix 1.

Because the data collected from the 30 replications of the factorial design (shown in Table A1.1. of Appendix 1) appears to be quite variable, it can be expected that the data from the pilot design will be even more so. The total number of customers in the pilot design is in fact only 9.375% of the total number of customers in the classical factorial design. Looking at Table 5.3., the average estimated values for $c(\gamma,p)$ appear reasonable, but the average estimated values for $\upsilon^2(\gamma,p)$ do not seem consistent with those found in Table 5.2. For example, the ratio of the average value of $\upsilon^2(\gamma,p)$ at the first design point to that at the fourth design point is out by a factor of $^1/_2$.

In particular, one of the two marginal cost function estimates at $\gamma = 0.95$ for pilot experiments 1, 11, 26 and 30 (see Table A1.2. in Appendix 1) is extremely low. For pilot experiment 26, the cost at (0.95, 0.75) is almost the same as at (0.8, 0.25) and (0.8, 0.75). As a result, it is likely that the designs found using these cost functions will perform badly.

After obtaining the data from the pilot experiment, we now need to determine a model for the marginal cost function $c(\gamma,p)\upsilon^2(\gamma,p)$. A simple model is a series of 4 triangular planes fitted over portions of the design region, joining 3 design points as shown in Figure 5.6.

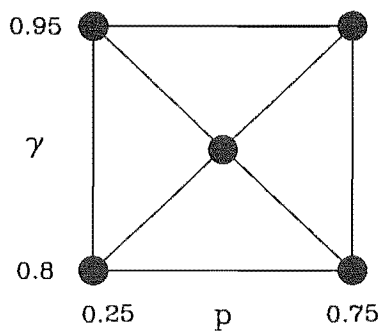


Figure 5.6. Fitting the marginal cost function

Using the data from Table A1.2. of Appendix 1 we can put together the 30 marginal cost functions associated with the 30 pilot experiments. The marginal cost function for the values shown in Table 5.3. is shown in Figure 5.7.

As mentioned before, for the design criterion target we will use the average value of the design criterion resulting from the 30 replications of the factorial design, 0.265. In general, the experimenter would be required to determine an appropriate value for this. The data collected during the pilot experiment can be a guide, as it provides rough estimates of the mean value of the response at various points in the design region. This allows a sensible value for the desired variance of these values to be determined.

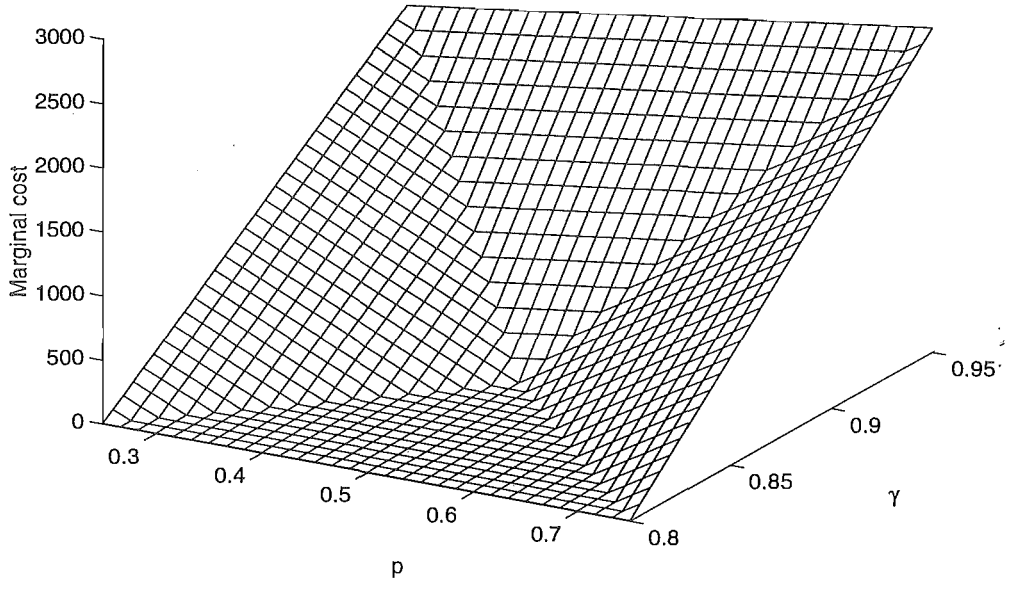


Figure 5.7. Marginal cost function for Table 5.3.

The SICOED design problem then becomes:

$$\text{Min} \quad \sum_{i=1}^{r} \frac{c(\gamma_i,p_i)\upsilon^2(\gamma_i,p_i)}{\sigma_i^2}$$

$$\text{s.t.} \quad \int_{0.8}^{0.95}\int_{0.25}^{0.75} \mathbf{f}^T(\gamma,p)\left(\sum_{i=1}^{r}\frac{1}{\sigma_i^2}\mathbf{f}(\gamma_i,p_i)\mathbf{f}^T(\gamma_i,p_i)\right)^{-1}\mathbf{f}(\gamma,p)\partial p\partial\gamma \Big/ \int_{0.8}^{0.95}\int_{0.25}^{0.75}\partial p\partial\gamma \le 0.265$$

$$0.8 \le \gamma_i \le 0.95 \quad \forall i$$

$$0.25 \le p_i \le 0.75 \quad \forall i$$

$$\sigma_i^2 \ge 0 \quad \forall i$$

where $\mathbf{f}(\gamma,p) = [1/(1-\gamma)\ \ \gamma/(1-\gamma)\ \ p/(1-\gamma)]^T$, and $c(\gamma,p)\upsilon^2(\gamma,p)$ is replaced by the appropriate marginal cost function. This design problem was then solved for each of the 30 marginal cost functions, as well as for the mean marginal cost function values shown in Table 5.3. The solution method used was the 3-phase heuristic method described in section 10 of Chapter 4, with an 11x11 grid over the design region. The design for the mean cost function values, which took 47.62 seconds to find, is shown in Table 5.4.

γ	p	$\sigma^2(\gamma,p)$
0.8	0.25	0.0408
0.8	0.75	0.0299
0.875	0.5	0.0543

Table 5.4. SICOED Design using Means of Cost Function Estimates

The main differences between the SICOED design in Table 5.4. and the factorial design in Table 5.2. are:

- The SICOED design has 3 design points, while the factorial design has 4.
- The factorial design requires experiments at the very costly setting $\gamma = 0.95$, while the highest value of γ in the SICOED design is 0.875.
- Comparing the values of Mean $\hat{V}ar(\overline{W}(\gamma,p))$ for the factorial design with the σ_i^2 for the SICOED design, we see that the latter design shifts some of the experimental effort away from the lower marginal cost region, where additional experiments result in a relatively small amount of information, to the higher marginal cost region, where additional experiments result in a relatively large amount of information.

The designs for the 30 individual marginal cost functions, and the time taken to find them, can be found in Table A1.3. of Appendix 1. The design points in (γ,p) space are graphed in Figure 5.8. Note that these graphs show p on the horizontal axis, γ on the vertical axis, and the required $\sigma^2(\gamma,p)$ to the right.

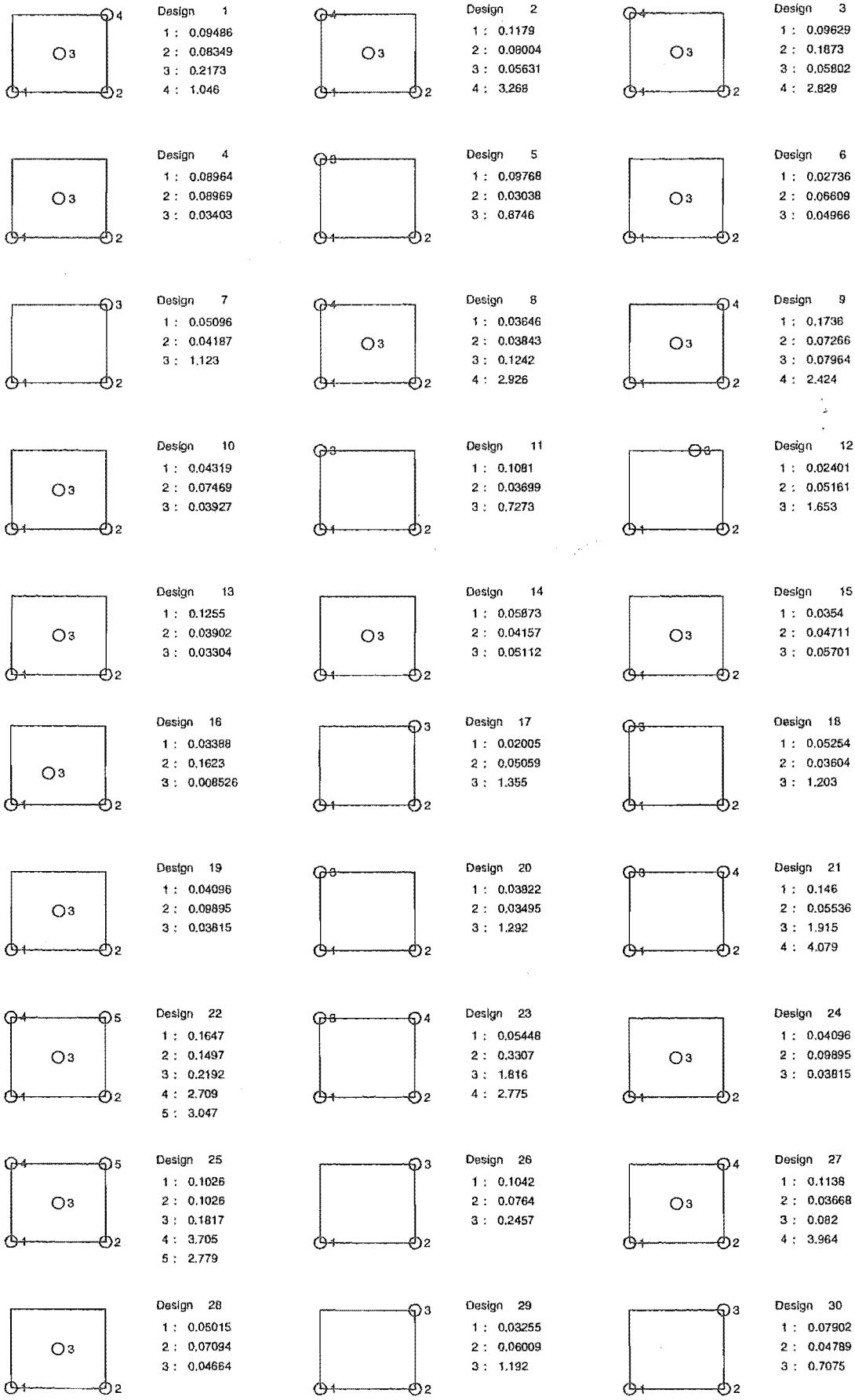


Figure 5.8. Designs using the SICOED approach

Because there was a large variability in the marginal cost function estimates, the designs resulting from the SICOED approach are also quite varied. Of the 30 designs, 20 are 3-point designs, 8 are 4-point designs, and 2 are 5-point designs.

The simulation model used to test these designs is similar to the model used to test the classical design, in that each run consists of 20,000 customers with a warm-up period of 10,000 customers. However, instead of automatically performing 20 such runs at each design point, the model was modified to include a simple stopping condition on s_i^2, the estimator of $\mathrm{Var}(\overline{W}(\gamma,p))$:

$$n_i = \min\{n:\ n \geq 5,\ (s_i^2 / n) \leq \sigma_i^2\}.$$

For each point, an initial sample of 5 runs was performed. Another run would then be performed at the current design point as long as the estimated variance of the mean response was larger than the value of $\sigma^2(\gamma,p)$ in the design.

The results for 30 replications of the 'mean' design in Table 5.4. are shown in Table 5.5.

γ	p	$\sigma^2(\gamma,p)$	*Mean* s_i^2	*Mean Runs per Replication*	*Total Cost (sec)*
0.8	0.25	0.0408	0.0313	8.07	5030
0.8	0.75	0.0299	0.0223	7.43	4081
0.875	0.5	0.0543	0.0513	30.20	19780

Table 5.5. Simulation results for the 'mean' cost function (30 replications)

As noted before, the optimal design resulting from the SICOED approach shifts some of the experimental effort from the low-cost end of the design region to the high-cost end (compared to the factorial design). This is because an additional run at a low value of γ (where the variance of the response is low) results in only a very small reduction in the design criterion value, whereas an additional run at a high value of γ (where the variance of the response is high) results in a much

larger reduction in the design criterion value. Overall, this shift allows the same design criterion value to be reached with fewer runs.

The simulation results of the 30 designs from the SICOED approach (shown in Figure 5.8.) are shown in Table A1.3. of Appendix 1.

Conclusions

The optimal design in Table 5.5., although not based on a very accurate marginal cost function, still has a total cost for 30 replications that is 52.6% lower than the total cost of 30 replications of the classical factorial design. When the cost of the pilot experiment and design algorithm is included, this figure becomes 41.2%. Also, the average design criterion value of the responses collected for the SICOED design is 0.228, significantly lower than the corresponding value for the classical design of 0.265.

However, for a number of reasons that will be discussed shortly, the results of the individual simulations shown in Table A1.3. of Appendix 1 do not appear as promising as the results of the 'mean' design would suggest. Figure 5.9. presents a summary of the individual simulation results for both the 30 repetitions of the classical design, and the 30 designs obtained using the SICOED approach.

Figure 5.9. shows the impact of using very inaccurate marginal cost function approximations on the experimental cost of the designs found using the SICOED approach. One-third of the SICOED designs have an experimental cost that is below the cost of the factorial designs. However, a further 17 designs have a cost that is between 5 and 100% more than the factorial designs, and there are a further 3 outliers. These outliers have a very large cost, and are the result of very extreme errors in the marginal cost function estimates used to generate those designs. For example, we commented earlier that the marginal cost function used for design number 26 was extremely inaccurate, and the result is that the experimental cost for this design was close to 7000 seconds.

On the other hand, despite the errors in the marginal cost function estimates, the design criterion values (based on the data collected) for the designs from the SICOED approach tend to lie well below the value of 0.265 required by the

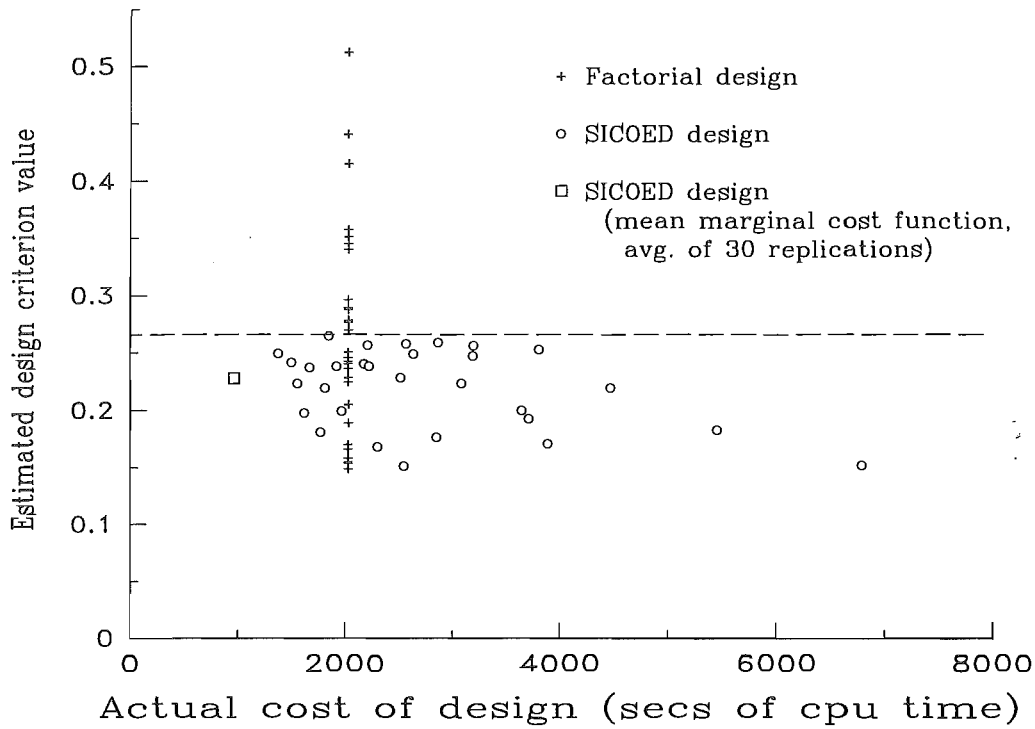


Figure 5.9. Comparison of classical and SICOED designs

design problem. The classical design resulted in a fairly large range of design criterion values, from 0.15 to 0.51, but each design required the same experimental cost.

These observations show one of the basic differences between the SICOED approach and the classical approach. The former ensures that a given level of information is achieved, while the latter ensures a fixed experimental cost (although note that generally only an estimate of this cost is known before experimentation).

However there are also a number of factors specific to this example that impact on the above observations.

- The size of the pilot experiment, or amount of prior information available, has a large impact on the results. A larger pilot experiment (consisting of more runs and / or longer runs) would lead to better and less variable marginal cost function estimates. In turn, this would improve the efficiency of the optimal

design, and thus reduce the expected experimental cost as well as the variability of this cost. In our Jackson Network example, the design for the pilot experiment requires only 3 pieces of data to be collected at each design point, resulting in highly variable marginal cost function estimates.

- As the overall size of the experiment increases, the expected mean percentage difference between the cost of the SICOED approach and the cost of the classical approach will stay the same (assuming the marginal cost function estimates stay constant). However the more runs are performed, the less variable the cost of the SICOED approach will be. Also, an increase in the overall size of the experiment allows a larger pilot experiment to be performed, resulting in an additional reduction in cost variability.
- Figure 5.9. is shaped to a large extent by the use of independent replications. First, the effect of an integer number of run means that the designs for the SICOED approach result in more data being collected than was necessary. This is due to the fact that while the optimal $\sigma^2(\gamma,p)$ are taken to be continuous, the discrete nature of simulation runs means that s_i^2 is a step function of the number of experiments performed. So when the stopping condition is reached, slightly more data has been collected than necessary. Hence many of the little circles in Figure 5.9. lie well below the dashed line. This effect reduces as the overall size of the experiment is increased, however the use of a variance estimation technique such as spectral analysis would also substantially reduce this effect. Second, the use of independent replications means that only 3 (mean) responses were collected at each point of the pilot design, meaning that the estimate of the variance function was based on the variance of 3 numbers. Again, the use of a technique like Spectral Analysis would allow more accurate estimates to be made for the same experimental cost.
- Finally, a simple limit on the number of experiments performed at any design point would prevent very badly estimated marginal cost functions from resulting in a very large experimental cost. When the number of experiments reaches this limit, it indicates that the marginal cost function should be re-estimated using the data collected. For this example, the restriction that $n_i \leq 50$ would have removed many of the outliers seen in Figure 5.9.

One issue that has not been discussed for this example is the problem of bias. In section 4 of Chapter 4 we outlined the two sources of bias in the design criterion estimate, resulting from the use of estimators that were used in the example in this section. In the next section we will use a Monte Carlo experiment to examine this bias.

5.4. Estimating the Design Criterion Value Using the Actual Data Collected: A Monte Carlo Study of Bias

As discussed in Chapter 3, the main focus of the SICOED approach is to ensure that the value of the design criterion meets (or falls slightly below) a specified target L_0. Thus if the design criterion and its target are selected to represent the experimental objective, then the experimenter can be sure that this objective is reached. For the SICOED approach, as for the classical design approaches, the process of experimental design is completely separate from the process of collecting the data. Because of this separation, the value of the design criterion based on the optimal experimental design E^* (labelled $L(E^*)$) will generally be different from the *estimate* of the design criterion value based on the *actual* data collected (labelled $L(\mathbf{x}_i, s_i^2)$).

First, consider the experimental design phase. We have assumed that $L(E^*)$ is a function of the metamodel parameter covariance matrix, and that the covariance matrix is found using Weighted Least Squares. Since we know both the design points $\mathbf{x}_i$ and the target variance σ_i^2 exactly, then $L(E^*)$ is not an estimate but an exact value. Unless additional constraints have been added to the standard SICOED design problem, any optimisation procedure used to find E^* will normally ensure that $L(E^*)$ is nearly exactly equal to L_0.

Once the design E^* has been found, experiments are performed sequentially until the stopping rule

$$n_i = \left\{ \min n: n \geq n_0, s_i^2(n) \leq \sigma_i^2 \right\}$$

is satisfied, where $s_i^2(n) = \hat{var}(y_i)/n$. The variance of the mean response $var(\bar{y}_i|n_i)$ for each design point is then estimated by $s_i^2(n_i)$. In turn, the mean response variance estimates are used to estimate the value of the design criterion $L(\mathbf{x}_i, s_i^2)$. As seen in the previous section, the sequential component of the SICOED approach ensures that $s_i^2 \leq \sigma_i^2$, and thus $L(\mathbf{x}_i, s_i^2) \leq L(E^*) \approx L_0$.

However, the estimator $L(\mathbf{x}_i, s_i^2)$ provides a biased estimate of $L(\mathbf{x}_i, var(\bar{y}_i))$, as discussed in section 4 of Chapter 4. First, the value of s_i^2 resulting from the sequential sample-size selection procedure shown above underestimates the true variance $var(\bar{y}_i|n_i)$. Second, the Estimated Weighted Least Squares estimator of the metamodel parameter covariance matrix (upon which most design criteria are based) is also biased, even if s_i^2 was unbiased. The result is that $L(\mathbf{x}_i, s_i^2)$ underestimates $L(\mathbf{x}_i, var(\bar{y}_i))$.

In this section we use a Monte Carlo simulation experiment to investigate two important questions. First, given that $L(\mathbf{x}_i, s_i^2)$ is a biased estimate of $L(\mathbf{x}_i, var(\bar{y}_i))$, how large is this bias and what influences it? Second, since $L(\mathbf{x}_i, s_i^2) \leq L(E^*)$ (as seen in the previous section), then is $L(\mathbf{x}_i, var(\bar{y}_i)) \leq L(E^*) \approx L_0$? The answers to these questions will provide an indication of whether or not the bias identified above should be a cause for concern, and how it may be reduced.

Details of the Inputs

In this Monte Carlo experiment we simulate the full process of experimentation. First we find an experimental design using the SICOED approach, by finding a solution to the design problem using the 3-phase heuristic developed in Chapter 4. We then 'perform' the required experiments by sampling from a known response distribution. The response data collected is then analysed by estimating the design criterion value.

The first step is to find an experimental design. To complete the specification of the design problem, we need to select the design criterion, its

target L_0, the number of factors considered, the form of the metamodel, the design region $\mathcal{X}$, and the marginal cost function. The inputs to the Monte Carlo experiment for this stage of the experimental process were as follows:

- *Design criterion*: As in the examples in sections 1 to 3, we study the 'average variance of the mean response' criterion. The behaviour of this criterion is likely to be similar to other criteria based on the variance of the mean response.
- *Design criterion target*: This value influences the total number of experiments to be performed. However, we use another input (α, discussed shortly) to control this, and have set $L_0 = 0.02$.
- *Number of factors*: We study a one-factor (x_1) and a two-factor (x_1, x_2) situation.
- *Form of metamodel*: We study a full first-order polynomial metamodel and a full second order polynomial metamodel.
- *Design region*: This is chosen to be $\mathcal{X} = \{x_1: 0 \le x_1 \le 1\}$ for the one-factor situation, and $\mathcal{X} = \{x_1, x_2: 0 \le x_1 \le 1, 0 \le x_2 \le 1\}$ for the two-factor situation.
- *Marginal cost function*: We set the cost-per-experiment function to be constant, and study the following variance functions:

$$\begin{aligned}
\text{Flat:} \quad & \upsilon^2(x) = 0.01 \\
\text{Medium:} \quad & \upsilon^2(x) = 0.01 + x_1/20 \qquad [+x_2/20] \\
\text{Steep:} \quad & \upsilon^2(x) = 0.01 + x_1 \qquad [+x_2] \\
\text{U-shape:} \quad & \upsilon^2(x) = 0.01 + (x_1 - 0.5)^2 \qquad [+(x_2 - 0.5)^2]
\end{aligned}$$

where the component in square brackets is added only in the case of a two-factor situation. These variance functions (which are also the marginal cost functions) are labelled as Flat, Medium, Steep and U-shape respectively.

The 3-phase heuristic developed in Chapter 4 was used to determine an experimental design for each of the 16 different experimental situations (= 4 combinations of number of factors and model order, and 4 variance functions). An 11-point equally spaced grid was used for each variable, leading to 11

candidate design points for the one-factor situations, and 11x11 = 121 candidate design points for the two-factor situations.

The second step is to simulate the actual experimentation. We do this by using a know distribution to generate the responses, rather than an actual simulation model such as the Jackson queueing network model in section 3. The advantages of this approach are that it is faster (no need for lengthy simulation runs), ensures that there is no specification bias in the metamodel form or marginal cost function, and that the bias in the estimate of the design criterion value can be calculated exactly. Since we are only interested in the variability of the responses, and not the responses themselves, we simply generate the 'random error' for each response. For this Monte Carlo experiment, we use the Normal distribution to generate errors with mean zero and variance equal to the variance functions shown above. This seems reasonable as the means of simulation runs (the mean responses) are often approximately Normally distributed.

We sample from the response distribution in the sequential manner discussed in the introduction to this section. An initial sample of size n_0 is taken, and the estimated variance of the mean response

$$s_i^2(n_0) = \frac{1}{n_0} \sum_{j=1}^{n_0} (y_{ij} - \bar{y}_i)^2$$

is calculated. If $s_i^2(n_0)$ is greater than the target variance σ_i^2, further samples are sequentially taken until $s_i^2(n_i) \leq \sigma_i^2$. A number of studies (see section 4 of Chapter 4) have investigated the minimum sample size required to obtain approximately asymptotic results for the EWLS estimators, and found this to be about 25 to 30. Hence in this Monte Carlo experiment we study three values of n_0: 4, 10 and 25.

We also introduce another input into our Monte Carlo experiment, to study the effect of the sequential sample-size selection procedure described above. We know that such a procedure leads to a biased estimator of the variance of the response mean. When $s_i^2(n_0)$ underestimates $\text{var}(\bar{y}_i | n_0)$, there is a high chance that sampling stops prematurely, while if it overestimates there is a high chance that sampling is continued and a more accurate estimate obtained. On average, $s_i^2(n_i)$ underestimates $\text{var}(\bar{y}_i | n_i)$. Logically, it would appear that this bias is related

to the difference between n_0, and the 'correct' sample-size $\bar{n} = \text{var}(y_i)/\sigma_i^2$. If $\bar{n} \leq n_0$, then the average sample size resulting from the sequential procedure will be close to n_0. This is similar to a non-sequential procedure with sample size n_0, and thus results in little bias. Similarly, when $\bar{n} >> n_0$ then there will be very few occurrences where the sequential procedure stops at a low sample size due to an underestimated mean response variance. Again, the bias is small. This suggests that the bias in s_i^2 will be largest when $\bar{n}$ is slightly greater than n_0.

The value of $\bar{n}$ depends directly on the size of the variance of the responses. Hence we can investigate the effect discussed above, by selecting different values for that variance. Since the relative values of the variance function should not change (so that the design problem does not change), we simply scale the variance function appropriately. The procedure we use to do this is as follows: For each design, we find a multiplier for the variance function such that the average (over the design points) of $\bar{n}$ is equal to $\alpha * n_0$ (we need to average over the design points because each design point may have a different σ_i^2). In the Monte Carlo experiment, we study the values α = 0.5, 1, 1.5, 2, 4, and 6. Note that since the sample size of each experiment is largely a function of α, we do not study different values of L_0, the design criterion target.

Of course, the sequential sample size procedure does still take at least n_0 samples at each design point. From the Monte-Carlo results, a rough guide to the effect of α on the actual sample size is that α values of 0.5, 1, 1.5, 2, 4, and 6 lead to actual sample sizes of approximately 1.05, 1.15-1.20, 1.5, 1.8-1.9, 4, and 6 times n_0.

To summarise, the inputs to the Monte Carlo experiment are the number of factors (1 or 2), the order of the metamodel (first or second), the cost function (flat, medium, steep or U-shape), the minimum sample size n_0 (4, 10 or 25) and the sample-size factor α (0.5, 1, 1.5, 2, 4, or 6). Thus a total of 2*2*4*3*6 = 288 situations are studied.

The final step of the process is to calculate the design criterion estimate $L(\mathbf{x}_i, s_i^2)$, which is found by evaluating the design criterion using the design points $\mathbf{x}_i$ and mean response variance estimates s_i^2.

The analysis of the output of the Monte Carlo experiment is as follows. First we calculate the unbiased estimate of the design criterion based on the actual sample size, $L(x_i, var(\bar{y}_i))$, which is found by using the actual sample size at each design point to calculate the theoretical variance of the response mean $var(\bar{y}_i)$. Then, rather than estimate the absolute value of the bias in the design criterion estimator $L(x_i, s_i^2)$, we consider the proportion

$$b_q = \frac{L_q(x_i, var(\bar{y}_i)) - E\left[L_q(x_i, s_i^2)\right]}{L_q(x_i, var(\bar{y}_i))},$$

where the q = 1, ..., 288 is used to label the different experimental situations investigated. As noted earlier, $L(x_i, s_i^2)$ is an underestimate of $L(x_i, var(\bar{y}_i))$, and hence b lies between 0 and 1. We repeat each of the 288 situations 250 times to reduce the effect of random variation in the results, and estimate b by

$$\hat{b}_q = \frac{L_q(x_i, var(\bar{y}_i)) - \frac{1}{250}\sum_{j=1}^{250} L_{qj}(x_i, s_i^2)}{L_q(x_i, var(\bar{y}_i))}.$$

Due to random variation, the value of $\hat{b}_q$ may be negative, although on average we would expect it to be positive. We also estimate the variance of $\hat{b}$ using the usual estimator

$$var(\hat{b}_q) = \frac{\sum_{j=1}^{250}\left(\hat{b}_{qj} - \hat{b}_q\right)^2}{250}.$$

Analysis of Monte Carlo Outputs

The experimental design for each combination of number of factors, model order and variance function, as well a table showing the associated value $\hat{b}$ (and its variance in brackets) for each combination of n_0 and α, are shown in Appendix 2.

First, we will briefly discuss the experimental designs. The designs very clearly show the effect of the different variance functions used. For the Flat variance function, the designs are (nearly) symmetrical, while the designs for the Medium and Steep variance functions are skewed and heavily skewed respectively. The designs for the U-shape variance function are also (nearly) symmetrical, but the target variance at the centre is relatively significantly smaller than for the Flat variance functions (note that the σ_i^2 values shown are for the design problem with $L_0 = 0.02$).

Since we are interested in estimating both the size of the bias as measured by $\hat{b}$, as well as the influence of the inputs to the Monte Carlo experiment on this bias, it appears appropriate to fit a regression model to the data. The values of $\hat{b}$ shown in Appendix 2 are independent, and approximately normally distributed (being averages of 250 samples), so the usual regression assumptions hold. Because we have estimates of their variances, we use Estimated Weighted Least Squares (EWLS) with $1/\text{var}(\hat{b}_q)$ as the weights. Due to the large sample size for each combination of inputs (250), the EWLS estimators should be very nearly unbiased.

We selected the terms in the regression model as follows. We use a 0-1 dummy variable for the number of factors and model order, and a further 3 dummy variables to represent the different variance function choices. Since we would expect that $\hat{b}$ is an asymptotically decreasing function of n_0, with the asymptote at zero, we include the term $1/n_0$. Lastly, as explained above we expect $\hat{b}$ to initially increase, stabilise, and then decrease as a function of α. Hence we include a cubic polynomial in α.

The resulting regression model is

$$\hat{b} = \lambda_1 + \lambda_2 \text{FACT} + \lambda_3 \text{MODEL} + \lambda_4 \text{VAR1} + \lambda_5 \text{VAR2} + \lambda_6 \text{VAR3} + \frac{\lambda_7}{n_0} + \lambda_8 \alpha + \lambda_9 \alpha^2 + \lambda_{10} \alpha^3,$$

where FACT = 0 for one-factor model
1 for two-factor model
MODEL = 0 for first-order model
1 for second-order model

VARx = 0 if that variance function is not used

1 if that variance function is used

(x: 1 = Flat, 2 = Medium, 3 = Steep)

The SAS/STAT report, showing the parameter estimates as well as various goodness-of-fit measures for the regression model, is shown in Figure 5.10.

Model: MODEL1
Dependent Variable: B

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	0.04007	0.00445	212.982	0.0001
Error	278	0.00581	0.00002		
C Total	287	0.04588			

Root MSE	0.00457	R-square	0.8733
Dep Mean	0.17164	Adj R-sq	0.8692
C.V.	2.66383		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-0.179668	0.01240341	-14.485	0.0001
FACT	1	0.040556	0.00505701	8.020	0.0001
MODEL	1	0.026236	0.00495053	5.300	0.0001
VAR1	1	0.003547	0.00693677	0.511	0.6096
VAR2	1	0.007005	0.00700017	1.001	0.3178
VAR3	1	-0.002882	0.00686991	-0.419	0.6752
INVN0	1	1.089149	0.02767616	39.353	0.0001
ALPHA	1	0.200999	0.01536551	13.081	0.0001
ALPHASQ	1	-0.061073	0.00581881	-10.496	0.0001
ALPHACUB	1	0.005180	0.00060963	8.497	0.0001

Figure 5.10. Fitting the regression model: Output from SAS/STAT

The F value shown in Figure 5.10. indicates that the overall regression is highly significant. The t-statistics for the parameters shows that except for the variance function dummy variables, in each case the parameter is significantly different from zero.

Because there are no interaction terms in the model, the parameters of the regression model are fairly easily interpreted. On average, $\hat{b}$, the proportion of bias in the design criterion estimate relative to its true value, increases by 0.04 when there are two factors rather than one. Similarly, $\hat{b}$ increases by 0.026 when

the model is a second-order model rather than a first-order model. This would suggest that bias increases as the number of factors and model order increases, although further Monte Carlo experiments are needed to confirm this. It is also possible that the important variable is the number of parameters in the metamodel.

Interestingly, the variance functions due not appear to have a significant impact on the proportional bias. Certainly the separate t-tests of the significance of the parameters show that λ_4, λ_5 and λ_6 are not significantly different from zero (see Figure 5.10). We also perform an F-test for the joint hypothesis that $\lambda_4 = \lambda_5 = \lambda_6$. The test statistic is 0.7617, while the F-value at the 1% level with 2 (= number of independent conditions implied by the hypothesis, being $\lambda_4 = \lambda_5$ and $\lambda_5 = \lambda_6$) and 278 (= 288 - number of parameters in full model) degrees of freedom is 4.61. Hence we accept the null hypothesis, that the variance function has no impact on $\hat{b}$.

The initial sample size used by the sequential procedure, n_0, is inversely related to bias. At $n_0 = 4$, the proportional bias due to n_0 is 0.27, while at $n_0 = 25$ it is only 0.04.

The input α is related to the proportional bias by a cubic polynomial function. A table of several α values, and the additional proportional bias resulting from those values, is shown below.

α	0.5	1	1.5	2	4	6
Additional bias	0.085	0.144	0.180	0.197	0.156	0.123

From these results, we can make a number of conclusions about the size of the proportional bias, and suggestions as to how it can be kept to a minimum. First, the maximum bias for the situations investigated here occurs when the experimental situation has 2 factors, the metamodel is a second-order polynomial, $n_0 = 4$ and $\alpha \approx 2.5$. For this situation, $\hat{b} = 0.36$ (using the mean of λ_3 to λ_6 as the intercept), which in most cases would be unacceptable. However, simply by increasing n_0 to 10, this figure substantially decreases to 0.20. At the other end of the scale, the minimum bias for the situations investigated occurs when there is only one factor, the metamodel is a first-order polynomial, $n_0 = 25$ and $\alpha = 0.5$.

Using the regression model, $\hat{b} = -0.04$ for this situation, but in practice we would expect a $\hat{b}$ that was at or close to zero. Interestingly, the shape of the variance function does not appear to have any effect on bias.

Generally, the experimenter has little control over the number of factors and the metamodel order, and in any case it would probably be unwise to change these simply to reduce the bias in the design criterion value estimator. When the simulation run-lengths are fixed in some way (e.g. terminating simulations), the variance of the responses is also fixed, and thus we are only able to influence α by adjusting n_0. Using the above results, we would recommend that n_0 be set as large as possible, *provided the resulting value of* α *is either high (>4) or low (<1)*. A value of $n_0 = 10$ would appear to reduce the proportional bias to a reasonable level of no more than 0.20 for the situations investigated. In general, however, we cannot assume that a larger value of n_0 will lead to less bias, as this will also have an effect on the value of α.

When the simulation run-lengths must also be selected as part of the experimental process (e.g. independent replications in steady-state simulation), then the issue of minimising bias becomes more complicated. We then are able to set α by (i) setting n_0, and (ii) indirectly by setting the variance of the responses through selecting the run-length. Providing guidelines for this case is difficult, because the run-length will in turn impact on the efficiency of the variance estimation method used. In any case, either during experimentation or after all runs have been completed, an estimate of the number of experiments performed at each design point n_i can be made. If $n_i / n_0 < 1.5$, bias is likely to be small; while for $n_i / n_0 \geq 1.5$. the value of α may be approximated by n_i / n_0, and the previous conclusions about α used to assess the bias.

We now use further results obtained from the Monte Carlo experiment to provide an indication of the answer to the second question posed at the start of this section: If $L(\bar{x}_i, s_i^2) < L(E^*)$, but $L(\bar{x}_i, s_i^2)$ is biased, then is $L(\bar{x}_i, \text{var}(\bar{y}_i)) \leq L(E^*) \approx L_0$? This is an important question, because the answer determines whether the SICOED approach using the biased design criterion value estimator (for lack of an alternative) does indeed ensure that the design criterion target is

met. If this is the case, it might be possible to use L_0 as an estimator for $L(x_i, var(\bar{y}_i))$.

Figure 5.11. shows a histogram of the 288 values of

$$\frac{L_0 - L_q(x_i, var(\bar{y}_i))}{L_0},$$

calculated from the Monte Carlo results.

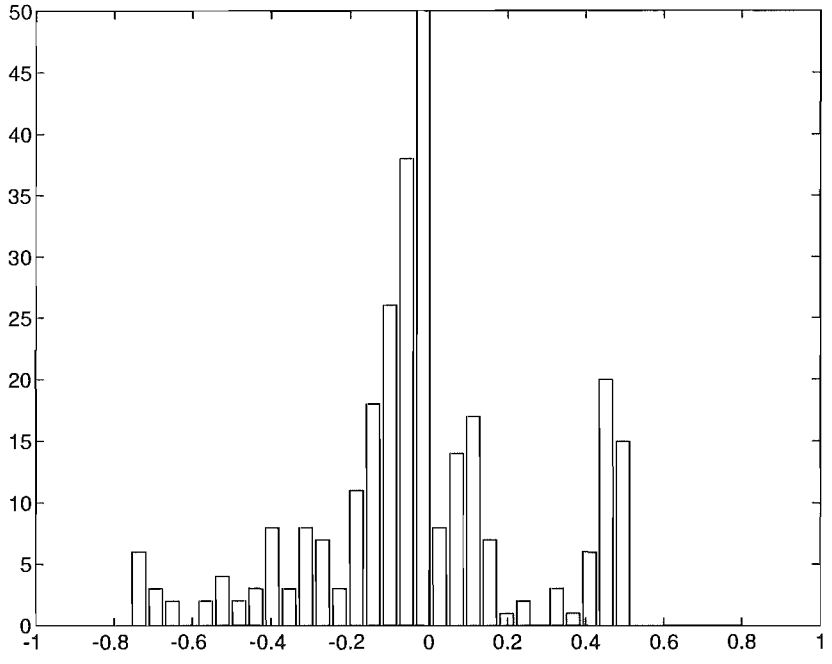


Figure 5.11. Histogram of $\left(L_0 - L_q(x_i, var(\bar{y}_i))\right)/L_0$

In the histogram, a negative value on the horizontal axis means that $L_q(x_i, var(\bar{y}_i)) > L_0$, while a positive value indicates that $L_q(x_i, var(\bar{y}_i)) < L_0$. For the design criterion target to be met or exceeded, we need all, or at least the majority, of the bars in the histogram to lie to the right of zero. However, nearly the opposite appears to be the case. Of the 288 input combinations studied in the Monte Carlo experiment, only 99 have a $L_q(x_i, var(\bar{y}_i))$ value that is equal to, or less than, L_0. The spread of bars along the horizontal axis of the histogram is also much larger to the left of zero than to the right. For some cases studied, the value of $L(x_i, var(\bar{y}_i))$ was nearly 75% higher than L_0.

However, further analysis shows that the results are better than they seem:

- In approximately 65% of the cases, $L(\mathbf{x}_i, \text{var}(\bar{y}_i))$ lies within 20% of L_0.
- In each case where $L(\mathbf{x}_i, \text{var}(\bar{y}_i))$ is more than 20% smaller than L_0, the value of n_0 was 4.
- In each case when $L(\mathbf{x}_i, \text{var}(\bar{y}_i))$ is more than 40% larger than L_0, the value of α was 0.5.

If we exclude from the 288 experiments any experiment where $n_0 = 4$ or $\alpha = 0.5$, then we find that the value of $\left(L_0 - L_q(\mathbf{x}_i, \text{var}(\bar{y}_i))\right)/L_0$ for the remaining 160 experiments has a mean of -0.03 and range (-0.198, 0.156). So provided we set $n_0 \geq 10$, and $\alpha \geq 1$, then in every one of the remaining cases examined in the Monte Carlo experiment we have $L(\mathbf{x}_i, \text{var}(\bar{y}_i))$ within 20% of L_0. This means that the main objective of the SICOED approach - to ensure that $L(\mathbf{x}_i, \text{var}(\bar{y}_i)) \leq L_0$ - is reasonably well satisfied for those cases, and that L_0 might be a more reasonable estimator of $L(\mathbf{x}_i, \text{var}(\bar{y}_i))$ than $L(\mathbf{x}_i, s_i^2)$.

5.5. Summary

In this chapter, we have studied a number of numerical examples in order to provide some indication of the behaviour of the SICOED approach. Three particular examples have been used to illustrate certain points made in previous chapters, and a Monte Carlo experiment has been used to indicate the effect of biased estimators.

In section 1, we illustrated a point made in Chapters 2 and 3: That the distribution of information (measured by the variance of the mean response at each point in the design region) resulting from the classical optimal design approach is likely to be different from that expected. This is the case in any situation where the variance function is not known exactly. On the other hand this is not the case for the SICOED approach, except that in most cases more 'information' than required will be collected due to the sequential element of the approach.

In section 2 we used another example, simulation of the M|M|1 queue, to provide an indication of the differences between the experimental design and associated experimental cost for each of the existing design approaches identified in Chapter 2. The differences in experimental cost between the approaches were found to be substantial. However these differences depend strongly on the cost and variance functions for the particular example used. In general, the experimental cost for the SICOED approach will not be much less than for classical designs (which implicitly or explicitly assume constant variance and cost) if the variance function and cost function are nearly constant. This difference becomes larger when there are large differences between the values of these functions across the design region.

One conclusion that can be drawn from the examples considered in sections 2 and 3 is that the rough shape of the estimated marginal cost function has a large influence on the efficiency of the SICOED approach. In section 2, the SICOED approach was insensitive to the exact slope of the estimated marginal cost function *provided* it was an increasing function of the factor. Similarly in section 3, the SICOED designs that performed very badly correspond to marginal cost function estimates that were a decreasing function of the arrival rate γ. Most simulation practitioners would have some idea of the likely shape of the marginal cost function for their simulation model. In the Jackson Network example, estimates of this function were found and used without modification. Some of the resulting marginal cost functions had a negative slope in the direction of γ, yet brief consideration of the simulation model shows that instead, the variance of W increases steeply with γ. Not surprisingly, the resulting designs did not perform well. In practice, practitioners would be advised to carefully consider the estimates of the marginal cost function used, and modify them if necessary.

At two stages in the Jackson Network example, an estimator is used that is known to underestimate. These are the estimator of the variance of the mean response at each design point s_i^2, and the Estimated Weighted Least Squares estimator of the metamodel parameter covariance matrix. The joint effect of these

estimators is that the design criterion value is underestimated. At this time, no alternative unbiased estimators appear to exist.

In section 4 we investigated the size of the bias in the estimator of the design criterion value. We studied 288 different experimental situations, consisting of one or two factors, a first or second order metamodel, 4 different variance functions, 3 values of n_0, and 6 values of a parameter α (related to the ratio of actual sample size n_i to n_0). A regression function was fitted to the data obtained, which showed that the largest effect on bias was from the variables n_0 and α. Surprisingly, an F-test on the relevant regression parameters showed that for the experimental situations studied, bias does not appear to depend on the shape of the variance function.

The largest amount of bias in the estimate of the design criterion value, as a percentage of the true value, was found to be 33%. When we restrict n_0 to be 10 or greater, this became approximately 20%. We also investigated the relationship between the true design criterion value and its target L_0, with a view to the possibility of using the latter as an estimator of the former. Again, large differences were found, but this time these differences were both positive and negative. However if we use the restrictions $n_0 \geq 10$ and $\alpha \geq 1$, then the difference as a percentage of L_0 was found to be no more than 20%.

It is difficult to provide practical guidelines for keeping bias to a minimum. On the one hand, n_0 should be as large as possible but at least 10, while on the other, we should have $1.5 \leq n_i/n_0 \geq 4$. These are likely to be opposing in many situations. However, it does appear that L_0 is a better and certainly more conservative estimator of the design criterion value than $L(\mathbf{x}_i, s_i^2)$.

CHAPTER 6: SEQUENTIAL EXPERIMENTAL DESIGN

6.1. Introduction

The main aim of the research presented in this thesis has been to develop an experimental design approach that is suitable specifically for the simulation context. As outlined in Chapter 2, there are a number of differences between the simulation and classical contexts, including the validity of a number of assumptions. However from a practical point of view, the most important difference is that simulation experiments are performed on computers. Most optimal experimental design methods are also necessarily computer based, due to the number of calculations required. Hence in simulation, the processes of designing and performing experiments can be coupled much more closely and easily than in other contexts.

The first step in this direction has been the development of computer software that aids the experimenter in selecting the design, allows rapid evaluation of the efficiency of potential designs, and prepares a design output file that can be used as input by the simulation software used. References to papers reporting the development of such software were given in Chapters 1 and 2.

However such software simply automates routine tasks, and does little to (i) reduce the knowledge of experimental design theory required of the experimenter, (ii) reduce the number of decisions that the experimenter must make, and (iii) take advantage of the fact that simulation is computer based. But this is not simply because these considerations have been ignored. As argued in Chapter 2, we believe that classical experimental design theory is simply not suitable for implementation into experimental design software.

In response we have developed an alternative design approach, "Semi-sequential Information Constrained Optimal Experimental Design" (SICOED), specifically for the simulation context. Our approach overcomes a number of

shortcomings of the classical approach, and is relatively easily applied in practice. It also goes some way to take advantage of the computer based nature of simulation. First, the experimental design for our approach is found by solving an optimisation problem, which uses the speed of computer based optimisation software to overcome the problem of selecting the design. Second, our approach includes an element of sequentiality. In classical contexts, sequential experimentation based on stopping rules is often difficult or impossible, whereas this is easily implemented in the simulation context.

Although the SICOED approach goes some way to take advantage of the computer based nature of simulation, it does not take full advantage of the interaction possible between the processes of experimental design and simulation. In this final chapter we will discuss an approach that does do that: Sequential experimental design. Although sequential design is not a complex procedure, and is reasonably simple to implement, there are a number of research issues that must be resolved before it can be confidently used in practice.

In section 2 we outline the main limitations of the SICOED approach. In section 3 we discuss the advantages and possible formats of a fully sequential design procedure, and in section 4 we identify some of the research issues surrounding sequential design.

6.2. Limitations of the SICOED Approach

The main limitations of the SICOED approach stem from the fact that it is only semi-sequential. The experimental design consists of stopping rules, which means that the total sample-size at each design point is to some extent influenced by the variance function. But the process of finding the design is strictly separated from the process of simulation. This leads to the following limitations:

1. We cannot use design criteria that depend on the values of the response, because in practice these are unknown at the time that the experimental design is determined. This prevents the use of design criteria that express relative,

rather than absolute, objectives. For example, such a criterion could be the half-width of the confidence interval of the fitted response, as a percentage of the fitted response (both averaged over the design region). The value of the design criterion is then simply a percentage, which means that an appropriate value for L_0 is more easily selected by the experimenter.

2. The efficiency of the design depends on the accuracy of the marginal cost function estimate. This function is part of the input to the experimental design process, and is not updated during experimentation, even if the original estimate was very inaccurate.
3. The SICOED approach (like the classical optimal approach) assumes that the form of the metamodel is known, and that any discrepancy between the simulation responses and the fitted metamodel is due to variance error only. If the latter is not the case, the design will not be efficient and invalid inferences may be made from the metamodel.
4. As discussed in section 3 of Chapter 5, the discrete nature of simulation runs means that more data is collected at each design point than is necessary. As a result, the design criterion value based on the responses obtained may be significantly lower than L_0. Our approach currently does not allow the target value of σ_i^2, for the design points at which experimentation has not yet been started, to be changed based on the response data already obtained.

For each of these points, there are ways to reduce the effect of the limitation implied by them. To overcome point 1, a pilot experiment can be used to provide an estimate of the mean response, and thus allow a sensible value of L_0 to be selected. Although points 2 and 4 indicate that our approach may be inefficient in some situations, the main objective of collecting sufficient information will be met regardless of the efficiency of the design. By allowing multiple metamodel forms, as shown in section 5 of Chapter 4, the effect of point 3 can be reduced.

Nevertheless, each of these limitations could be improved on.

6.3. Format and Advantages of a Sequential Design Approach

Sequential design is the process of selecting the experimental design sequentially during experimentation, rather than before any experimentation takes place. At each stage of the process, the data collected from previous stages is used to determine which experiments are to be performed in the next stage. The process terminates when the design criterion target, or some other stopping condition, is met. A sequential design procedure takes into account the actual experimental situation encountered, and is not limited to the initial estimates or guesses provided by the experimenter.

Sequential design can take many forms, depending on which aspects of the design problem are allowed to change when new information is obtained during experimentation. Three examples of this discussed in this thesis are:

- Simple design replication: Repetitions of a design are carried out until a stopping condition is reached (see section 3 of Chapter 3). Hence only the overall sample size N is sequentially determined.
- Semi-sequential: The design points are fixed, but stopping rules on the variance of the mean response at each point determine n_i (this is the SICOED approach).
- Fully sequential: Here all the components of the design problem are re-selected or re-estimated during experimentation at least once, and a new design for the remaining experiments determined. This includes the form of the metamodel, the marginal cost function, and potentially the design region, design criterion and its target.

We will use the term 'sequential design' to mean the last of these examples. Besides the components of the design problem that could be updated during experimentation, there is the question of how frequently such an update is done. Generally a 2-stage or multi-stage procedure is used, such as those commonly seen in sequential analysis.

A multi-stage sequential design procedure requires frequent re-selection and re-estimation of various components of the design problem, after which the new design problem must be solved to find the new design. Computer based simulation is thus the ideal context for sequential design, as these tasks can be performed by experimental design software that interacts with the simulation software. A diagram showing the various processes and how these processes could be connected is shown in Figure 6.1.

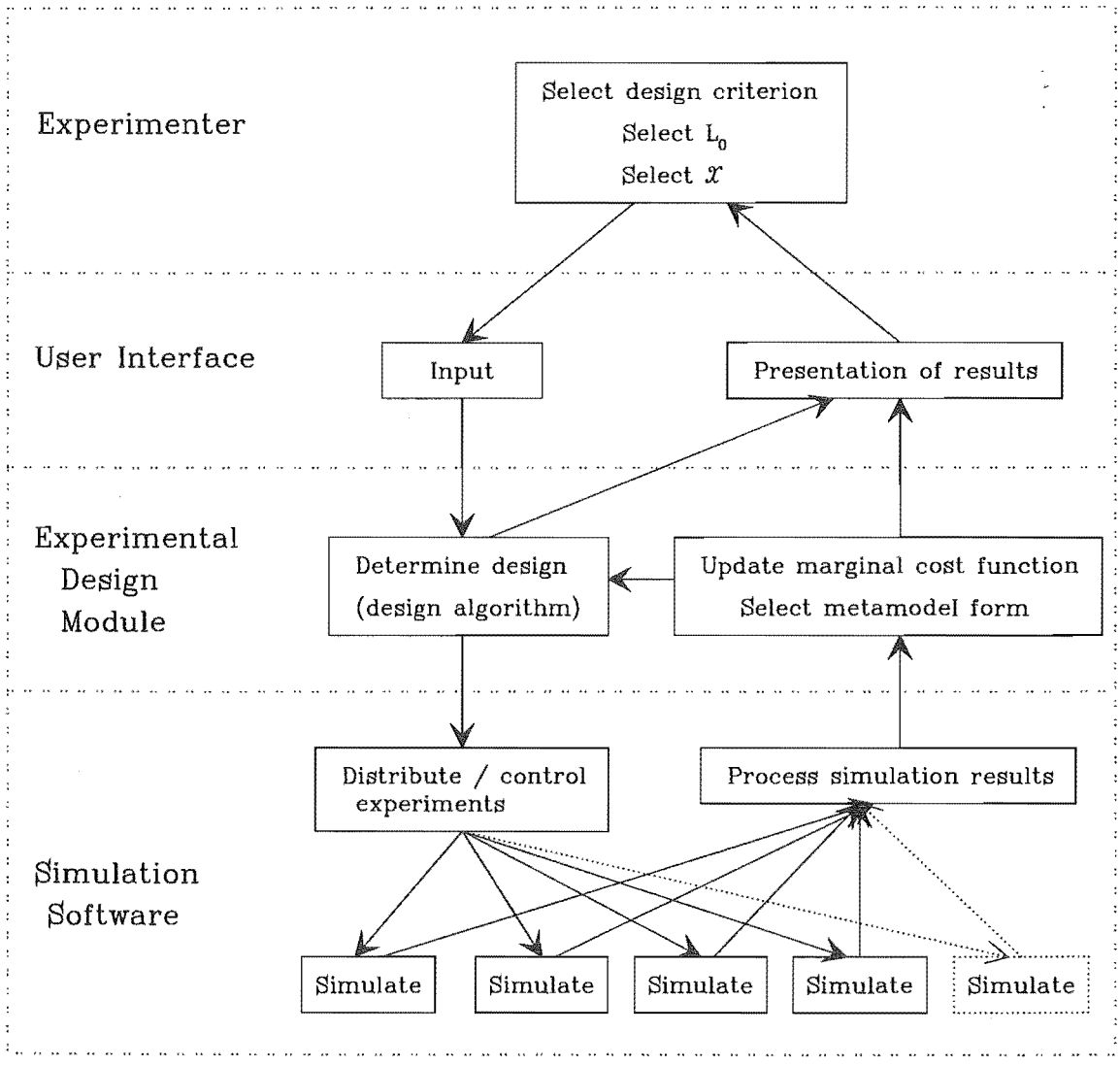


Figure 6.1. Sequential design: Illustration of processes and their interactions

Figure 6.1. assumes that the simulation software used is able to distribute 'copies' of the simulation program, with different parameter settings, onto various computers connected to a network, such as the simulation package AKAROA

(see Pawlikowski, Yau, and McNickle (1994)) can. In such a case, the design problem would need to be modified to ensure that the different cost of experimentation on each computer is taken into account.

A small number of researchers have made detailed suggestions for sequential design procedures in classical contexts (e.g. Fedorov (1972), Sokolov (1963a,b)), but again this does not appear to have been the subject of much research in the simulation literature. One exception is Donohue, Houck and Myers (1993b), who present a sequential method, in 4 stages, using classical design-property designs (see section 3 of Chapter 3). Their sequential approach is not concerned with sample size, which is assumed to be set by the experimenter.

The two main advantages of a sequential design approach are a potential increase in efficiency, and reduced demands on the experimenter. The main reason for an increase in efficiency is that better estimates of the marginal cost function are obtained as experimentation progresses, leading to a more efficient design. As seen in section 3 of Chapter 5, the marginal cost function estimate has a very substantial impact on the efficiency of the design. Also, during a sequential procedure the data collected may indicate that the form of the metamodel is different from the initial model specified. Indeed we may add to the design problem the requirement that sufficient data be collected to allow such hypotheses to be tested. The result is that a more accurate metamodel form is selected.

Sequential analysis also has a number of practical advantages, which reduce the demands for (often unknown) information from the experimenter. First, because data from previous stages is used to determine the design for the next stage, estimates of the mean response can be made. This allows the use of design criteria that express relative, rather than absolute, values. Setting an appropriate target for such 'relative criteria' is significantly simpler, as they can be expressed as percentages. Also, the initial selections of the marginal cost function and the form of the metamodel become less important, as these can be updated during the sequential design process. Together, this means that sequential design does not

assume that the experimenter has much knowledge of the experimental situation prior to experimentation.

6.4. Some Research Issues for Sequential Design

It is clear that a sequential design approach has a number of significant advantages, and it would appear reasonably straightforward to implement a procedure such as that outlined in Figure 6.1. into software. However there are a number of issues that need to be investigated before it is possible to advise the use of sequential design.

First, sequential design would appear to be most suitable for large simulation studies, where a significant amount of data is collected. For smaller studies the minimum amount of data required by the procedure to establish estimates of the mean response (for a 'relative' criterion), metamodel form, and / or marginal cost function, may be too great. Research is needed to establish guidelines for when a sequential procedure can reasonably be applied.

Second, there are a number of issues regarding the details of any sequential design procedure. For example, should the design be expressed in terms of n_i or σ_i^2? The advantage of the former is that the 'natural' size of each stage is a discrete number of runs. However, this definition of an experiment has the disadvantages identified in Chapter 2, such as the inability to use variance estimation methods other than Independent Replications. When the design is expressed in terms of σ_i^2, and (say) Spectral Analysis is used to estimate the variance of the mean response, it is possible to stop at (nearly) any stage of a run, rather than being forced to finish each run. But in turn, this requires a decision to be made about the definition of a stage, as this can no longer consist simply of a certain number of runs (as there may only be one run at each design point).

Probably the most important research issue in sequential design is the issue of bias. As discussed in section 4 of Chapter 5, the usual estimators of variance

(and even mean) are often biased when applied to data collected using a sequential analysis procedure. For example, when a stopping rule based on variance is used to select the sample size at a design point, the usual estimator of the mean response variance is biased. In sequential design, we also adaptively choose the design points, marginal cost function, and even the form of the metamodel. The Monte Carlo study in section 4 of Chapter 5 showed that in some cases there may be substantial bias in the estimator of the design criterion value when using the SICOED approach. It is possible that a fully sequential design approach will lead to an increase in the size of this bias. The effect of using such an approach on the usual estimator for the design criterion value does not appear to have been investigated, even by those who have suggested sequential design procedures (Fedorov (1972), Sokolov (1963a,b), and Donohue, Houck and Myers (1993b) make no mention of the validity of the usual estimators). It is possible that the amount of bias is unacceptable, in which case new estimators may need to be developed.

6.5. Summary

Computer simulation is the ideal context for which to consider the application of a sequential experimental design approach. Such an approach has a number of advantages over the semi-sequential SICOED approach. Most importantly, a fully sequential design approach would further reduce the amount of information required from the experimenter.

However, the transition from the non-sequential classical approach to the semi-sequential SICOED approach led to the introduction of some bias in the usual estimator of the design criterion value. Further sequentialisation of the design process may worsen this bias. Currently, no research into this question appears to have been reported. Before any particular sequential design procedure can be recommended, research is required to establish the resulting level of bias (if any) in the usual estimators.

SUMMARY AND CONCLUSION

In this thesis we have investigated experimental design methods for simulation. In particular, we have concentrated on stochastic simulation studies, where we wish to find an analytical model (referred to as a 'metamodel') relating the response of the simulation model with a set of factors.

In Chapter 2 we critically examined the literature on experimental design for simulation. In nearly all of this literature, classical designs like factorial and composite simplex designs are used and recommended. Classical designs were originally developed for the agricultural context, and it is not immediately clear that these designs are also suitable for the simulation context. However, there appears to have been very little discussion of the differences between the contexts.

We believe that a number of substantial differences do exist. Most importantly, classical design methods assume that the total number of experiments to be performed is known in advance, due to cost and / or time considerations. For simulation, the cost and time taken to perform experiments is not such an important consideration as in the classical context. Hence the emphasis often lies not on performing a given number of experiments, but rather on the amount of information obtained from those experiments. In addition, classical design methods generally assume that both the variance of the response, and the cost per experiment, are constant for every setting of the factors. This is often not the case in simulation. Finally, the application of classical design methods to simulation experiments requires the use of the Independent Replications method of estimating response variance, and leads to a number of inflexibilities.

Since simulation is computer based, it is the ideal context for which to develop experimental design software that could aid the experimenter in the design process. However, nearly every step of the selection of a classical design requires arbitrary decisions to be made. Hence software based on such designs

would do little more than automate a number of routine tasks. We conclude that a new experimental design approach for simulation should be developed.

In Chapter 3 we first investigated two existing alternatives to the design methods employed in the simulation literature. These are optimal experimental design theory, and sequential analysis. We also proposed a design approach consisting of the combination of these methods. However, although these methods have some advantages over the application of classical designs, they do not overcome some of the main concerns outlined above.

In the remainder of this chapter, we proposed a new approach to experimental design for simulation. Rather than assume that the total number of experiments to be performed is known exactly, the objective of our approach is to minimise the experimental cost while ensuring that a given level of information (or knowledge goal) is reached. The function measuring this level of information is known as the design criterion. Since our approach is based on the classical optimal design approach, the design for our approach is found by solving an optimisation problem, known as the design problem. In this way, the problem of choosing an experimental design is modelled through an optimisation problem, as opposed to arbitrarily selecting a design.

Rather than focus on the number of experiments performed, we focus on the variance of the mean response obtained from those experiments. Thus the experimental design consists of a set of experimental design points with associated stopping rules. We label this approach as "Semi-sequential, Information Constrained, Optimal Experimental Design", or SICOED.

The SICOED approach has a number of advantages over the design approach that has been used in the simulation literature. First, the sequential element of our approach ensures that the experimenter's information goal (given the estimators chosen) is reached, regardless of the accuracy of the estimate of the variance of the mean response function. This is in contrast to the classical approaches, where the focus lies on performing a given number of experiments. Second, the SICOED approach is easily incorporated into experimental design software. Given a small number of inputs, such software is able to determine the complete experimental design. Selection of the inputs requires only elementary

knowledge of experimental design theory. Third, the SICOED approach can be used in conjunction with variance estimation methods other than Independent Replications. This allows more efficient methods to be used. Lastly, the SICOED approach explicitly recognises that the variance of the mean response and cost-per-experiment functions are generally not constant, and is generally very flexible.

In Chapter 4 we considered some of the components of the SICOED design problem, and discussed solution methods. A number of suggestions for estimating the cost-per-experiment and variance functions were made, and a number of suitable design criteria developed. We also discussed the presence of bias in the design criterion estimate based on the actual responses collected, due to the estimators used.

Two examples were used to illustrate that in general, the design criterion will not be a convex function. This implies that a local optimal solution (design) to the SICOED design problem may not be a global optimal solution. However, if we assume that the number of candidate design points is finite, then a modified SICOED design problem does have the required convexity properties.

Three solution methods were discussed: Algebraic solution methods, non-linear optimisation methods, and heuristics. Algebraic solution methods are problem-specific and cannot easily be integrated into experimental design software. Non-linear solution methods can take a substantial amount of computer time to solve the SICOED design problem, and cannot guarantee to find a global optimal solution. On the other hand, heuristic solution procedures applied to the modified design problem can quickly determine a close-to (globally) optimal design. Thus we have developed a 3-Phase solution heuristic for the SICOED design problem, which consists partially of the exchange and sequential design algorithms found in the design algorithm literature. This heuristic is easily incorporated into experimental design software.

In Chapter 5 we illustrated a number of properties of the SICOED approach through three examples and a Monte Carlo study. The first example illustrated the effect of the sequential component of our approach, which ensures that the

amount and / or distribution of information (however this is measured) obtained by the design is as expected. In the second example, we show the increase in efficiency possible by using the SICOED approach. Even when the variance function estimate is inaccurate, the SICOED approach can be significantly more efficient than other approaches found in the literature.

In the third example, simulation of a Jackson queueing network, we show the complete process of design, experimentation and analysis using both a factorial design, and the SICOED approach. This example showed that in general, the SICOED approach is more suitable for 'large' simulation studies, where it is expected that a substantial number of experiments will be performed.

Finally, we investigated the size of the bias in the design criterion estimate based on the responses collected, using a Monte Carlo study. This bias (on average leading to an underestimate) is the result of the combination of two biased estimators: The usual estimator of variance, applied to data collected from a sequential procedure, and the Estimated Weighted Least Squares estimator. The Monte Carlo study consisted of simulating the design, experimentation, and analysis phases of the experimental process, for 288 different situations. The size of the bias was estimated using a regression function fitted to the data. The variable that had the largest influence on bias was found to be the initial sample size n_0 of the sequential procedure. However, for the situations studied the bias in the design criterion estimate was no more than 20% when $n_0 \geq 10$. In general, it appears that bias can be kept to an acceptable level by ensuring that n_0 is as large as possible, provided the ratio of n_0 to the actual number of experiments performed is either close to 1, or above 4. Also, it was found that if these conditions are met, then the design criterion target may be a more suitable estimator of the design criterion value than the usual estimator.

To summarise, the SICOED approach has a number of advantages over the classical design approaches used in the simulation design literature, and is easily incorporated into experimental design software. Unlike classical design methods, our approach can be applied to a wide range of simulation situations, and combined with many simulation techniques. However, further research is

required to develop alternative sequential procedures and estimators to reduce the presence of bias in the estimate of the design criterion value.

In the final chapter we discussed sequential experimental design. The SICOED approach is semi-sequential, as the design criterion is determined before experimentation takes place, but consists of a set of stopping rules. A fully sequential design / experimentation procedure would have a number of advantages. Design criteria that are a function of the mean response could then be used, such as the average confidence interval width as a percentage of the mean response. Setting targets for such criteria would be significantly simpler. Also, a sequential procedure would allow the marginal cost function to be updated periodically, leading to a more efficient design.

However, for the SICOED approach the usual estimator of the design criterion value is a biased estimator, partially because of the semi-sequential nature of the approach. It would seem likely that a fully sequential design procedure would introduce further bias, unless unbiased estimators are derived.

The simulation context is ideally suited for the use of sequential experimental design methods, as computer-based design and experimentation procedures can be closely coupled. However sequential design has not received much attention in the literature. Further research is required to establish the extent of any bias, and develop appropriate sequential design procedures.

REFERENCES

Albert, A. (1966) "Fixed size confidence ellipsoids for linear regression parameters", *The Annals of Mathematical Statistics*, 37, 1602-1630.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*, Wiley, New York.

Atkinson, A.C. (1969) "Constrained maximisation and the design of experiments", *Technometrics*, 11, 616-618.

Atkinson, A.C. (1970) "The design of experiments to estimate the slope of a response surface", *Biometrika*, 57, 319-328.

Atkinson, A.C. (1982) "Developments in the design of experiments", *International Statistical Review*, 50, 161-177.

Atkinson, A.C. and Donev, A.N. (1992) *Optimum Experimental Designs*, Oxford University Press, Oxford.

Atkinson, A.C. and Cook, R.D. (1993) "D-optimum designs for heteroscedastic linear models", *Statistics Pre-Print Series #2*, London School of Economics.

Avriel, M. (1976) *Non-linear Programming: Analysis and Methods*, Prentice-Hall, N.J.

Barton, R. (1993) "New tools for simulation metamodels", *Working paper 93-110*, Industrial and Management Systems Engineering, Pensylvania State University.

Bazaraa, M.S. and Shetty, C.M. (1979) *Nonlinear Programming: Theory and Algorithms*, Wiley, New York.

Biles, W.E. (1974) "A gradient-regression search procedure for simulation experimentation", in *Proceedings of the 1974 Winter Simulation Conference*, H.J. Highland (ed.), 491-497.

Biles, W.E. and Swain, J.J. (1979) "Mathematical programming and the optimisation of computer simulations", in *Engineering Optimisation*, M. Avriel, R.S. Dembo (eds.), 189-207, North-Holland, Amsterdam.

Box, G.E.P. and Cox, D.R. (1964) "An analysis of transformations", *Journal of the Royal Statistical Society (Series B)*, 26, 211-252.

Box, G.E.P. and Draper, N.R. (1959) "A basis for the selection of a response surface design", *Journal of the American Statistical Association*, 54, 622-654.

Box, G.E.P. and Draper, N.R. (1963) "The choice of a second order rotatable design", *Biometrika*, 50, 335-352.

Box, G.E.P. and Draper, N.R. (1975) "Robust designs", *Biometrika*, 62, 347-352.

Box, G.E.P. and Draper, N.R. (1982) "Measures of lack of fit for response surface designs and predictor variable transformations", *Technometrics*, 24, 1-8.

Box, G.E.P. and Draper, N.R. (1987) *Empirical model-building and response surfaces*, John Wiley & Sons, New York.

Box, G.E.P. and Hunter, J.S. (1957) "Multi-factor experimental designs for exploring response surfaces", *Annals of Mathematical Statistics*, 28, 195-241.

Box, G.E.P. and Wilson, K.B. (1951) "On the experimental attainment of optimum conditions", *Journal of the Royal Statistical Society (Series B)*, 13, 1-45.

Box, M.J. and Draper, N.R. (1971) "Factorial designs, the |X'X| criterion, and some related matters", *Technometrics*, 13, 731-742.

Brooks, S.H. and Mickey, M.R. (1961) "Optimum estimation of gradient direction in steepest ascent experiments", *Biometrics*, 17, 48-56.

Chaloner, K. (1984) "Optimal Bayesian experimental design for linear models", *The Annals of Statistics*, 12, 283-300.

Cheng, R.C.H. and Kleijnen, J.P.C. (1995) "Optimal design of simulation experiments with nearly saturated queues", *Working Paper*, Institute of Mathematics and Statistics, University of Kent at Canterbury, United Kingdom.

Chernoff, H. (1972) *Sequential Analysis and Optimal Design*, SIAM, Philadelphia.

Cook, R.D. and Nachtsheim, C.J. (1980) "A comparison of algorithms for constructing exact D-optimal designs", *Technometrics*, 22, 315-324.

Deaton, M.L. (1983) "Estimation and hypothesis testing in regression in the presence of nonhomogeneous error variances", *Communication in Statistics - Simulation and Computation*, 12, 45-66.

Donohue, J.M. (1994) "Experimental design for simulation", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila (eds.), 200-206.

Donohue, J.M., Houck, E.C. and Myers, R.H. (1992) "Simulation designs for quadratic response surface models in the presence of model misspecification", *Management Science*, 38, 1765-1791.

Donohue, J.M., Houck, E.C. and Myers, R.H. (1993a) "Simulation designs and correlation induction for reducing second-order bias in first-order response surfaces", *Operations Research*, 41, 880-902.

Donohue, J.M., Houck, E.C. and Myers, R.H. (1993b) "A sequential experimental design procedure for the estimation of first- and second order simulation metamodels", *ACM Transactions on Modelling and Computer Simulation*, 3, 190-224.

Draper, N.R. (1963) "Ridge analysis of response surfaces", *Technometrics*, 5, 469-479.

Draper, N.R. and Smith, H. (1981) *Applied regression analysis (2nd ed.)*, Wiley, New York.

Dykstra Jr., O. (1971) "The augmentation of experimental data to maximise |X'X|", *Technometrics*, 13, 682-688.

Farrell, W. (1977) "Literature review and bibliography of simulation optimisation", in *Proceedings of the 1977 Winter Simulation Conference*, 117-124.

Fedorov, V.V. (1972) *Theory of optimal experiments*, Academic Press, New York.

Fisher, R.A. (1990) *Statistical Methods, Experimental Design and Scientific Inference*, Oxford University Press, Oxford.

Fishman, G.S. (1971) "Estimating sample size in computing simulation experiments", *Management Science*, 18, 21-38.

Fu, M.C. (1994) "A tutorial review of techniques for simulation optimisation", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila (eds.), 149-156.

Gardenier, T.K. (1990) "PRE-PRIM as a pre-processor to simulations: A cohesive unit", *Simulation*, 54, 65-70.

Ghosh, B.K. and Sen, P.K. (1991) (eds.) *Handbook of Sequential Analysis*, M. Dekker, New York.

Gleser, L.J. (1965) "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters", *The Annals of Mathematical Statistics*, 36, 463-467.

Grace, A. (1990) *Optimisation Toolbox for use with MATLAB*, The MathWorks, Inc.

Hader, R.J. and Park, S.H. (1978) "Slope-rotatable central composite designs", *Technometrics*, 20, 413-417.

Hartley, H.O. and Ruud, P.G. (1969) "Computer optimisation of second order response surface designs", in *Statistical Computing*, R.C. Milton, J.A. Nelder (eds), 441-462, Academic Press, New York

Hill, W.J. and Hunter, W.G. (1966) "A review of reponse surface methodology: A literature survey", *Technometrics*, 8, 571-590.

Hook, R. and Jeeves, T.A. (1961) " 'Direct search' solution of numerical and statistical problems", *Journal of the Association for Computing Machinery*, 8, 212-229.

Hossain, A. and Tobias, A. (1991) "WITNESS in the hands of an expert system: Using an expert system in the design and interpretation of simulation experiments", *OR Insight*, 4, 10-14.

Hunter, J.S. and Naylor, T.H. (1970) "Experimental designs for computer simulation experiments", *Management Science*, 16, 422-434.

Jacobson, S.H. and Schruben, L.W. (1989) "Techniques for simulation response optimisation", *Operations Research Letters*, 8, 1-9.

Joshi, S.S., Sherali, H.D. and Tew, J.D. (1994) "An enhanced RSM algorithm using gradient-deflection and second-order search strategies", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila (eds.), 297-304.

Karson, M.J., Manson, A.R. and Hader, R.J. (1969) "Minimum bias estimation and experimental design for response surfaces", *Technometrics*, 11, 461-475.

Khuri, A.I. and Cornell, J.A. (1987) *Response Surfaces*, Marcel Dekker, New York.

Kiefer, J. (1959) "Optimum experimental design", *Journal of the Royal Statistical Society (Series B)*, 21, 272-317.

Kiefer, J. and Wolfowitz, J. (1959) "Optimum designs in regression problems", *Annals of Mathematical Statistics*, 30, 271-292.

Kiefer, J. and Wolfowitz, J. (1960) "The equivalence of two extremum problems", *Canadian Journal of Mathematics*, 12, 363-366.

Kiessler, P.C. and Disney, R.L. (1982) "The sojourn time in a three node acyclic Jackson queueing network", *Technical Report VTR 8016*, Department of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University, Blacksburg.

Kleijnen, J.P.C. (1975) *Statistical Techniques in Simulation*, Part II, Marcel Dekker, New York.

Kleijnen, J.P.C. (1987) *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.

Kleijnen, J.P.C. (1992) "Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals", *Management Science*, 38, 1164-1185.

Kleijnen, J.P.C., Brent, R. and Brouwers, R. (1981) "Small-sample behaviour of weighted least squares in experimental design applications", *Communication in Statistics - Simulation and Computation*, 10, 303-313.

Kleijnen, J.P.C., Cremers, P. and van Belle, F. (1985) "The power of weighted and ordinary least squares with estimated unequal variances in experimental design", *Communication in Statistics - Simulation and Computation*, 14, 85-102.

Kleijnen, J.P.C., van den Burg, A.J. and van der Ham, R.T. (1979) "Generalization of simulation results - Practicality of statistical methods", *European Journal of Operational Research*, 3, 50-64.

Kleijnen, J.P.C., and van Groenendaal, W. (1992) *Simulation: A Statistical Perspective*, Wiley, Chicester.

Kleijnen, J.P.C. and van Groenendaal, W. (1995) "Two-stage versus sequential sample-size determination in regression analysis of simulation experiments", *Technical Report*, Dept of Information Systems and Auditing, School of Management and Economics, Tilburg University.

Lemoine, A. (1979) "On total sojourn times in networks of queues", *Management Science*, 25, 1034-1035.

Lindsey, J.K. (1972) "Fitting response surfaces with power transformations", *Applied Statistics*, 21, 234-247.

Mead, R. (1988) *The Design of Experiments*, Cambridge University Press, Cambridge.

Mead, R. and Pike, D.J. (1975) "A review of response surface methodology from a biometric viewpoint", *Biometrics*, 31, 803-851.

Meeker, W.Q. Jr., Hahn, G.J. and Feder, P.I. (1975) "A computer program for evaluating and comparing experimental designs and some applications", *The American Statistician*, 29, 60-64.

Meier, R.C. (1967) "The application of optimum-seeking techniques to simulation studies: A preliminary investigation", *Journal of Financial and Quantitative Analysis*, 2, 31-51.

Meidt, G.J. and Bauer, K.W. Jr. (1992) "PCRSM: A decision support system for simulation metamodel construction", *Simulation*, 59, 183-191.

Meketon, M.S. (1987) "Optimisation in simulation: A survey of recent results", in *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, H. Grant, W.D. Kelton (eds.), 58-67.

Mitchell, T.J. (1974) "An algorithm for the construction of D-optimal experimental designs", *Technometrics*, 16, 203-210.

Montgomery, D.C. and Evans, D.M. (1975) "Second order response surface designs in computer simulation", *Simulation*, 25, 169-178.

Murty, V.N. and Studden, W.J. (1972) "Optimal designs for estimating the slope of a polynomial regression", *Journal of the American Statistical Association*, 67, 869-873.

Myers, R.H. (1971) *Response Surface Methodology*, Allyn and Bacon, Boston.

Myers, R.H., Khuri, A.I. and Carter, W.H. Jr. (1989) "Response surface methodology: 1966-1988", *Technometrics*, 31, 137-157.

Myers, R.H. and Lahoda, S.J. (1975) "A generalisation of the response surface mean square error criterion with a specific application to the slope", *Technometrics*, 17, 481-486.

Nakayama, M.K. (1994) "Two-stage stopping procedures based on standardised time series", *Management Science*, 40, 1189-1206.

Neuhardt, J.B. and Bradley, H.E. (1971) "On the selection of multi-factor experimental arrangements with resource constraints", *Journal of the American Statistical Association*, 66, 618-621.

Nozari, A. (1984) "Generalized and ordinary least squares with estimated and unequal variances", *Communications in Statistics - Simulation and Computation*, 521-537.

Ott, L. and Mendenhall W. (1972) "Designs for estimating the slope of a second order linear model", *Technometrics*, 14, 341-353.

Park, S.H. (1987) "A class of mutifactor designs for estimating the slope of response surfaces", *Technometrics*, 29, 449-453.

Pawlikowski, K. (1990) "Steady-state simulation of queueing processes: A survey of problems and solutions", *ACM Computing Surveys*, 22, 123-170.

Pawlikowski, K., Yau, V.W.C. and McNickle, D. (1994) "Distributed stochastic discrete-event simulation in paralell time streams", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski, A.F. Seila (eds.), 723-730.

Pazman, A. (1986) *Foundations of Optimum Experimental Design*, Reidel, Dordrecht.

Pukelsheim, F. (1993) *Optimal Design of Experiments*, Wiley, New York.

Raktoe, B.L., Hedayat, A. and Federer, W.T. (1981) *Factorial Designs*, Wiley, New York.

Rawlings, J.O. (1988) *Applied Regression Analysis*, Wadsworth & Brooks, California.

Robertazzi, T.G. and Schwartz, S.C. (1989) "An accelerated sequential algorithm for producing D-optimal designs", *SIAM Journal of Scientific and Statistical Computing*, 10, 341-358.

Safizadeh, M.H. (1990) "Optimisation in simulation: Current issues and the future outlook", *Naval Research Logistics*, 37, 807-825.

Safizadeh, M.H. and Thornton, B.M. (1984) "Optimisation in simulation experiments using response surface methodology", *Computation & Industrial Engineering*, 8, 11-27.

Sanchez, P.J. (1994) (chair) "Simulation statistical software: An introspective appraisal", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski and A.F. Seila (eds.), 1311-1315.

Sargent, R.G. (1991) "Research issues in metamodeling", in *Proceedings of the 1991 Winter Simulation Conference*, B.L. Nelson, W.D. Kelton, G.M. Clark (eds.), 888-893.

Schmidt, P. (1976) *Econometrics*, Marcel Dekker, New York.

Schruben, L.W. and Margolin, B.H. (1978) "Psuedorandom number assignment in statistically designed simulation and distribution sampling experiments", *Journal of the American Statistical Association*, 73, 504-525.

Segreti, A.C., Carter Jr., W.H. and Wampler, G.W. (1979) "Monte carlo evaluation of several sequential optimisation techniques when the response is time to an event", *Journal of Statistical Computation and Simulation*, 9, 289-301.

Segreti, A.C., Carter Jr., W.H. and Wampler, G.W. (1981) "A monte carlo evaluation of the robustness of several sequential optimisation techniques when the response is time to an event", *Journal of Statistical Computation and Simulation*, 12, 209-216.

Silvey, S.D. (1980) *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Chapman and Hall, London.

Simon, B. and Foley, R.D. (1979) "Some results on sojourn times in acyclic Jackson networks", *Management Science*, 25, 1027-1034.

Smith, D.E. (1973a) "An empirical investigation of optimum-seeking in the computer simulation situation", *Operations Research*, 21, 475-497.

Smith, D.E. (1973b) "Requirements of an optimizer for computer simulations", *Naval Research Logistics Quarterly*, 20, 161-179.

Smith, D.E. (1976) "Automatic optimum-seeking program for digital simulation", *Simulation*, 27, 27-32.

Sokolov, S.N. (1963a) "Continuous planning of regression experiments I", *Theory of Probability and its Applications*, 8, 89-96.

Sokolov, S.N. (1963b) "Continuous planning of regression experiments II", *Theory of Probability and its Applications*, 8, 298-304.

Srivastava, M.S. (1967) "On fixed-width confidence bounds for regression parameters and mean vector", *Journal of the Royal Statistical Society (Series B)*, 29, 132-140.

Srivastava, M.S. (1971) "On fixed-width confidence bounds for regression parameters", *The Annals of Mathematical Statistics*, 42, 1403-1411.

Steinberg, D.M. and Hunter, W.G. (1984) "Experimental design: Review and comment", *Technometrics*, 26, 71-97.

Tao, Y. and Nelson, B.L. (1994) "A conceptual framework for simulation experiment design and analysis", in *Proceedings of the 1994 Winter Simulation Conference*, J.D. Tew, S. Manivannan, D.A. Sadowski, and A.F. Seila (eds.), 681-688.

Tew, J.D. and Wilson, J.R. (1992) "Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy", *Operations Research*, 40, 87-103.

van der Genugten, B.B. (1983) "The asymptotic behaviour of the estimated generalized least squares method in the linear regression model", *Statistica Neerlandica*, 37, 127-141.

van Meel, J.W., and Aris, J.D. (1993a) "'Cut and dried?' Generic instruments for regression based experimental design", in *European Simulation Symposium*, A. Verbraeck and E.J. Kerckhoffs (eds.), 553-558.

van Meel, J.W. and Aris, J.D. (1993b) "250 years of computer time? A simulation case study using regression based techniques for experimental design", in *European Simulation Symposium*, A. Verbraeck and E.J. Kerckhoffs (eds.), 35-40.

Welch, W.J. (1982) "Branch-and-bound search for experimental designs based on D optimality and other criteria", *Technometrics*, 24, 41-48.

Welch, P.D. (1990) "Simulation and regression: Dealing with the assumption of a common error variance", in *Proceedings of the 1990 Winter Simulation Conference*, O. Balci, R.P. Sadowski, R.F. Nance (eds.), 392-394.

Whitt, W. (1989) "Planning queueing simulations", *Management Science*, 35, 1341-1366.

Whitt, W. (1991) "The efficiency of one long run versus independent replications in steady state simulation", *Management Science*, 37, 645-666.

Wild, R.H. and Pignatiello, J.J. Jr. (1991) "An experimental design strategy for designing robust systems using discrete-event simulation", *Simulation*, 57, 358-368.

Wynn, H.P. (1970) "The sequential generation of D-optimum experimental designs", *Annals of Mathematical Statistics*, 41, 1655-1664.

Appendices

Appendix 1: Data from Jackson Network Example

Replication number	$\hat{\text{Var}}(\overline{W})$ at (γ, p)				Design criterion
	(0.8, 0.25)	(0.8, 0.75)	(0.95, 0.25)	(0.95, 0.75)	
1	0.0170	0.0083	4.1880	3.2319	0.2763
2	0.0281	0.0243	2.7627	11.5360	0.3571
3	0.0087	0.0104	4.1270	3.5949	0.2872
4	0.0086	0.0120	5.3632	2.6819	0.2695
5	0.0113	0.0062	3.0891	1.3869	0.1485
6	0.0161	0.0098	5.0676	4.3368	0.3507
7	0.0089	0.0084	2.9424	5.5872	0.2881
8	0.0114	0.0139	3.0249	2.3879	0.2049
9	0.0111	0.0216	2.3234	1.6388	0.1535
10	0.0075	0.0095	3.0896	4.6888	0.2781
11	0.0150	0.0191	4.9233	7.1989	0.4401
12	0.0178	0.0155	2.2453	4.8948	0.2403
13	0.0080	0.0122	2.2857	1.8697	0.1581
14	0.0148	0.0167	2.5784	3.4461	0.2282
15	0.0279	0.0179	8.7898	5.4705	0.5114
16	0.0168	0.0219	3.0785	3.2283	0.2456
17	0.0242	0.0168	4.1948	3.5138	0.2958
18	0.0171	0.0085	3.9985	5.4071	0.3447
19	0.0182	0.0142	4.2542	3.4061	0.2891
20	0.0145	0.0112	1.2685	6.2566	0.1702
21	0.0152	0.0226	2.4028	4.2897	0.2422
22	0.0169	0.0062	4.6702	2.1180	0.2245
23	0.0118	0.0061	1.8733	2.2108	0.1549
24	0.0131	0.0100	3.5862	2.1177	0.2044
25	0.0141	0.0155	3.4846	6.2955	0.3400
26	0.0178	0.0144	4.1711	8.0641	0.4146
27	0.0092	0.0167	3.2014	1.9705	0.1887
28	0.0123	0.0132	4.2753	2.4172	0.2363
29	0.0147	0.0091	2.6698	4.3264	0.2504
30	0.0172	0.0141	3.1055	1.5509	0.1658

Table A1.1. $\hat{\text{Var}}(\overline{W})$ **for 30 replications of classical factorial design**

Exp. number	Var($\overline{W}$) at (γ, p)					Cost (sec)
	(0.8, 0.25)	(0.8, 0.75)	(0.875, 0.5)	(0.95, 0.25)	(0.95, 0.75)	
1	7.300	7.164	14.881	2145.076	117.476	26.04
2	5.737	0.906	3.409	93.524	629.764	14.23
3	3.629	2.881	3.407	101.679	418.220	10.70
4	2.665	3.116	2.256	843.987	1093.379	43.12
5	3.521	0.198	17.358	74.823	863.189	40.54
6	1.617	5.651	17.654	1242.729	870.629	45.75
7	5.332	1.744	46.399	11090.720	263.209	9.00
8	6.832	4.305	93.877	2137.294	10151.550	11.27
9	4.990	3.522	7.936	608.850	171.845	11.53
10	0.647	5.964	4.186	2431.461	31197.150	35.27
11	13.466	5.193	237.392	128.630	10720.820	33.99
12	1.052	3.386	85.664	994.579	775.342	46.53
13	10.814	1.259	3.044	2691.080	649.655	40.03
14	2.056	0.861	7.353	1408.329	955.526	40.21
15	3.285	5.871	35.928	5980.016	1197.914	16.43
16	2.274	15.813	0.384	4695.685	545.216	44.38
17	4.312	15.393	146.548	4093.277	2001.827	11.36
18	5.268	3.721	79.335	485.153	2548.382	11.26
19	4.311	11.681	10.117	2895.290	4110.109	39.38
20	4.969	7.261	70.583	1042.091	2564.232	11.38
21	6.698	1.886	59.698	334.197	540.317	14.28
22	0.796	0.409	1.614	50.367	47.676	10.43
23	1.133	14.555	47.045	153.181	364.902	17.41
24	1.919	6.094	5.026	670.499	478.170	44.44
25	4.907	5.959	21.924	650.444	474.583	9.93
26	6.754	3.693	71.500	856.788	4.111	39.66
27	13.052	4.141	40.016	1281.609	1160.011	15.54
28	0.449	2.173	3.582	6578.673	339.159	39.77
29	1.576	3.907	42.989	705.153	292.572	10.93
30	5.586	2.184	55.844	1967.958	76.873	39.38

Table A1.2. Data from 30 pilot experiments, used to estimate the marginal cost function

Design number	γ_1	p_1	σ_1^2	$Var(\bar{y}_1)$ (Sim.)	γ_2	p_2	σ_2^2	$Var(\bar{y}_2)$ (Sim.)
1	0.800	0.250	0.095	0.0414	0.800	0.750	0.083	0.0818
2	0.800	0.250	0.118	0.1083	0.800	0.750	0.080	0.0751
3	0.800	0.250	0.096	0.0899	0.800	0.750	0.187	0.0496
4	0.800	0.250	0.090	0.0264	0.800	0.750	0.090	0.0103
5	0.800	0.250	0.098	0.0422	0.800	0.750	0.030	0.0257
6	0.800	0.250	0.027	0.0244	0.800	0.750	0.066	0.0549
7	0.800	0.250	0.051	0.0474	0.800	0.750	0.042	0.0204
8	0.800	0.250	0.036	0.0359	0.800	0.750	0.038	0.0350
9	0.800	0.250	0.174	0.0494	0.800	0.750	0.073	0.0115
10	0.800	0.250	0.043	0.0333	0.800	0.750	0.075	0.0730
11	0.800	0.250	0.108	0.0525	0.800	0.750	0.037	0.0282
12	0.800	0.250	0.024	0.0160	0.800	0.750	0.052	0.0298
13	0.800	0.250	0.125	0.0278	0.800	0.750	0.039	0.0383
14	0.800	0.250	0.059	0.0312	0.800	0.750	0.042	0.0409
15	0.800	0.250	0.035	0.0345	0.800	0.750	0.047	0.0442
16	0.800	0.250	0.034	0.0090	0.800	0.750	0.162	0.1113
17	0.800	0.250	0.020	0.0198	0.800	0.750	0.051	0.0415
18	0.800	0.250	0.053	0.0511	0.800	0.750	0.036	0.0343
19	0.800	0.250	0.041	0.0386	0.800	0.750	0.099	0.0919
20	0.800	0.250	0.038	0.0351	0.800	0.750	0.035	0.0337
21	0.800	0.250	0.146	0.1219	0.800	0.750	0.055	0.0492
22	0.800	0.250	0.165	0.0639	0.800	0.750	0.150	0.0407
23	0.800	0.250	0.054	0.0523	0.800	0.750	0.331	0.1163
24	0.800	0.250	0.041	0.0368	0.800	0.750	0.099	0.0821
25	0.800	0.250	0.103	0.0151	0.800	0.750	0.103	0.1026
26	0.800	0.250	0.104	0.0542	0.800	0.750	0.076	0.0568
27	0.800	0.250	0.114	0.0184	0.800	0.750	0.037	0.0358
28	0.800	0.250	0.050	0.0436	0.800	0.750	0.071	0.0526
29	0.800	0.250	0.033	0.0318	0.800	0.750	0.060	0.0498
30	0.800	0.250	0.079	0.0431	0.800	0.750	0.048	0.0214

Table A1.3. SICOED designs with simulation results

Design number	γ_3	p_3	σ_3^2	$Var(\bar{y}_3)$ (Sim.)	γ_4	p_4	σ_4^2	$Var(\bar{y}_4)$ (Sim.)
1	0.875	0.500	0.217	0.2122	0.950	0.750	1.046	1.0297
2	0.875	0.500	0.056	0.0539	0.950	0.250	3.268	3.1830
3	0.875	0.500	0.058	0.0420	0.950	0.250	2.829	2.7264
4	0.875	0.500	0.034	0.0338				
5	0.950	0.250	0.875	0.8642				
6	0.875	0.500	0.050	0.0485				
7	0.950	0.750	1.123	1.1157				
8	0.875	0.500	0.124	0.1218	0.950	0.250	2.926	2.8537
9	0.875	0.500	0.080	0.0796	0.950	0.750	2.424	2.2488
10	0.875	0.500	0.039	0.0388				
11	0.950	0.250	0.727	0.7174				
12	0.950	0.600	1.653	1.6117				
13	0.875	0.500	0.033	0.0322				
14	0.875	0.500	0.051	0.0502				
15	0.875	0.500	0.057	0.0561				
16	0.860	0.450	0.009	0.0090				
17	0.950	0.750	1.355	1.3410				
18	0.950	0.250	1.203	1.1770				
19	0.875	0.500	0.038	0.0377				
20	0.950	0.250	1.292	1.2867				
21	0.950	0.250	1.915	1.8903	0.950	0.750	4.079	3.7821
22	0.875	0.500	0.219	0.2126	0.950	0.250	2.709	2.6879
23	0.950	0.250	1.816	1.7968	0.950	0.750	2.775	2.6973
24	0.875	0.500	0.038	0.0374				
25	0.875	0.500	0.182	0.1497	0.950	0.250	3.705	3.5112
26	0.950	0.750	0.246	0.2455				
27	0.875	0.500	0.082	0.0799	0.950	0.750	3.964	3.4601
28	0.875	0.500	0.047	0.0460				
29	0.950	0.750	1.192	1.1586				
30	0.950	0.750	0.708	0.7043				

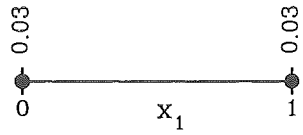
Table A1.3. SICOED designs with simulation results (cont...)

Design number	γ_5	p_5	σ_5^2	$Var(\bar{y}_4)$ (Sim.)	Heuristic cost (sec)	Sim. cost (sec)	Total runs	Design crit.
1					18.46	3083.95	90	0.2236
2					14.77	1383.28	52	0.2491
3					13.40	1970.70	56	0.1993
4					44.93	2554.04	98	0.1511
5					42.89	3712.35	101	0.1924
6					48.29	1673.66	67	0.2376
7					15.75	2183.07	67	0.2407
8					13.80	2214.68	72	0.2570
9					14.44	1776.65	59	0.1812
10					37.78	1559.81	62	0.2235
11					35.93	5452.19	142	0.1825
12					27.02	2567.77	72	0.2580
13					41.68	2307.20	91	0.1682
14					42.07	1927.73	76	0.2381
15					15.71	1849.59	74	0.2646
16					46.14	3891.02	156	0.1710
17					13.46	3805.49	107	0.2530
18					12.20	3199.49	89	0.2561
19					41.46	2230.23	88	0.2385
20					13.35	2867.25	81	0.2587
21					16.59	3194.24	89	0.2472
22	0.950	0.750	3.047	2.9732	12.52	3644.70	101	0.1994
23					17.52	4470.10	120	0.2191
24					46.24	2522.15	100	0.2283
25	0.950	0.750	2.779	2.6810	12.31	1621.02	47	0.1977
26					42.29	6790.60	187	0.1522
27					16.21	1817.64	64	0.2192
28					41.46	1501.78	60	0.2417
29					13.57	2640.89	81	0.2491
30					41.36	2855.55	81	0.1762

Table A1.3. SICOED designs with simulation results (cont...)

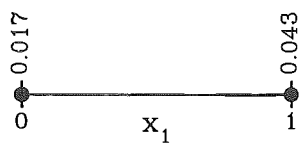
Appendix 2: Monte Carlo Results

Factors: 1 *Model order*: 1 *Variance function*: *Flat*



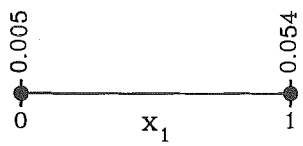
n_0 \ α	0.5	1	1.5	2	4	6
4	0.034 (0.028)	0.264 (0.020)	0.284 (0.021)	0.302 (0.020)	0.320 (0.021)	0.273 (0.022)
10	−0.018 (0.019)	0.098 (0.015)	0.157 (0.014)	0.172 (0.015)	0.104 (0.013)	0.076 (0.010)
25	−0.016 (0.013)	0.060 (0.011)	0.097 (0.011)	0.084 (0.010)	0.030 (0.007)	0.027 (0.006)

Factors: 1 *Model order*: 1 *Variance function*: *Medium*



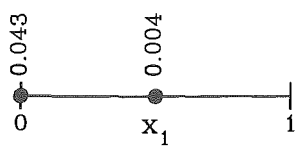
n_0 \ α	0.5	1	1.5	2	4	6
4	0.082 (0.030)	0.219 (0.025)	0.276 (0.024)	0.266 (0.022)	0.290 (0.020)	0.245 (0.019)
10	0.042 (0.020)	0.084 (0.020)	0.125 (0.017)	0.125 (0.016)	0.084 (0.011)	0.072 (0.011)
25	−0.003 (0.015)	0.077 (0.013)	0.044 (0.011)	0.048 0.009	0.049 (0.007)	0.022 (0.006)

Factors: 1 *Model order*: 1 *Variance function*: *Steep*



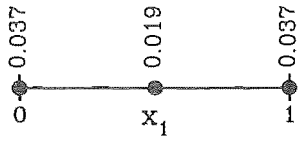
n_0 \ α	0.5	1	1.5	2	4	6
4	0.196 (0.033)	0.251 (0.029)	0.233 (0.029)	0.221 (0.027)	0.178 (0.023)	0.090 (0.017)
10	0.062 (0.023)	0.111 (0.020)	0.128 (0.018)	0.089 (0.016)	0.034 (0.011)	0.027 (0.009)
25	0.018 (0.015)	0.058 (0.013)	0.037 (0.010)	0.012 (0.010)	0.018 (0.007)	0.005 (0.005)

Factors: 1 *Model order*: 1 *Variance function*: *U−shape*



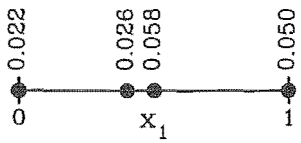
n_0 \ α	0.5	1	1.5	2	4	6
4	0.114 (0.032)	0.242 (0.025)	0.239 (0.023)	0.305 (0.022)	0.296 (0.021)	0.232 (0.019)
10	0.011 (0.022)	0.050 (0.019)	0.133 (0.017)	0.113 (0.015)	0.079 (0.012)	0.070 (0.010)
25	−0.003 (0.015)	0.067 (0.012)	0.038 (0.011)	0.068 0.009	0.029 (0.007)	0.023 (0.006)

Factors: 1 *Model order*: 2 *Variance function*: *Flat*



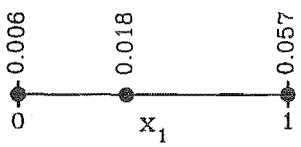
n_0 \ α	0.5	1	1.5	2	4	6
4	0.095 (0.026)	0.213 (0.021)	0.325 (0.018)	0.345 (0.018)	0.374 (0.017)	0.322 (0.017)
10	0.051 (0.018)	0.144 (0.015)	0.124 (0.013)	0.166 (0.013)	0.113 (0.011)	0.103 (0.009)
25	0.020 (0.012)	0.039 (0.010)	0.073 (0.009)	0.066 (0.008)	0.033 (0.006)	0.037 (0.005)

Factors: 1 *Model order*: 2 *Variance function*: *Medium*



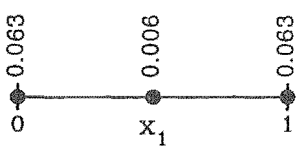
n_0 \ α	0.5	1	1.5	2	4	6
4	0.190 (0.022)	0.339 (0.017)	0.355 (0.017)	0.326 (0.017)	0.330 (0.017)	0.274 (0.016)
10	0.067 (0.014)	0.132 (0.013)	0.143 (0.013)	0.144 (0.012)	0.120 (0.009)	0.079 (0.008)
25	0.024 (0.009)	0.035 (0.009)	0.061 (0.008)	0.050 (0.008)	0.041 (0.005)	0.029 (0.004)

Factors: 1 *Model order*: 2 *Variance function*: *Steep*



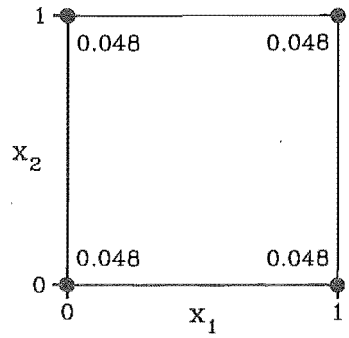
n_0 \ α	0.5	1	1.5	2	4	6
4	0.190 (0.023)	0.289 (0.023)	0.327 (0.021)	0.337 (0.022)	0.302 (0.021)	0.214 (0.021)
10	0.028 (0.018)	0.136 (0.014)	0.164 (0.015)	0.142 (0.015)	0.083 (0.011)	0.035 (0.008)
25	0.013 (0.012)	0.076 (0.011)	0.072 (0.010)	0.043 (0.008)	0.029 (0.005)	0.013 (0.004)

Factors: 1 *Model order*: 2 *Variance function*: *U−shape*



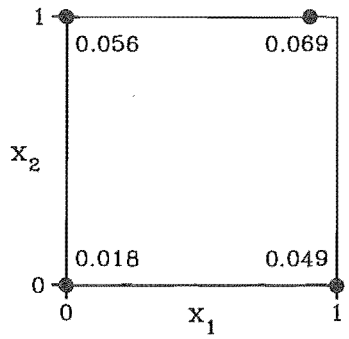
n_0 \ α	0.5	1	1.5	2	4	6
4	0.088 (0.023)	0.273 (0.018)	0.313 (0.017)	0.321 (0.018)	0.307 (0.018)	0.308 (0.019)
10	0.001 (0.017)	0.107 (0.013)	0.148 (0.014)	0.145 (0.014)	0.105 (0.011)	0.072 (0.009)
25	−0.011 (0.013)	0.047 (0.010)	0.065 (0.010)	0.051 (0.008)	0.036 (0.005)	0.023 (0.004)

Factors: 2 *Model order*: 1 *Variance function*: *Flat*



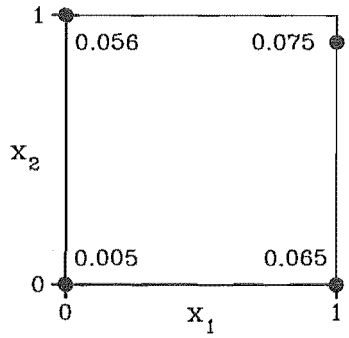
n_0 \ α	0.5	1	1.5	2	4	6
4	0.207 (0.020)	0.265 (0.016)	0.313 (0.016)	0.348 (0.015)	0.334 (0.014)	0.317 (0.015)
10	0.054 (0.014)	0.116 (0.012)	0.172 (0.010)	0.185 (0.010)	0.117 (0.009)	0.080 (0.007)
25	0.023 (0.008)	0.065 (0.008)	0.091 (0.007)	0.074 (0.007)	0.032 (0.005)	0.029 (0.004)

Factors: 2 *Model order*: 1 *Variance function*: *Medium*



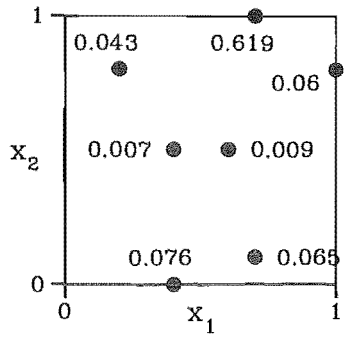
n_0 \ α	0.5	1	1.5	2	4	6
4	0.164 (0.021)	0.291 (0.018)	0.351 (0.016)	0.342 (0.017)	0.297 (0.015)	0.257 (0.015)
10	0.099 (0.014)	0.115 (0.014)	0.140 (0.011)	0.146 (0.011)	0.113 (0.009)	0.070 (0.007)
25	0.021 (0.010)	0.069 (0.008)	0.066 (0.008)	0.052 (0.006)	0.030 (0.005)	0.028 (0.004)

Factors: 2 *Model order*: 1 *Variance function*: *Steep*



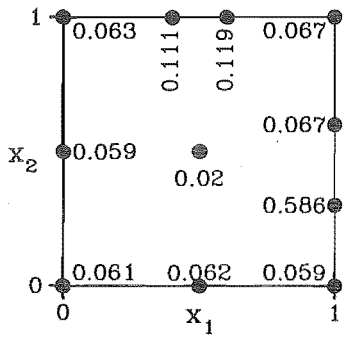
n_0 \ α	0.5	1	1.5	2	4	6
4	0.281 (0.020)	0.318 (0.018)	0.343 (0.019)	0.334 (0.017)	0.266 (0.016)	0.196 (0.016)
10	0.065 (0.015)	0.105 (0.013)	0.124 (0.013)	0.148 (0.012)	0.066 (0.008)	0.040 (0.007)
25	0.039 (0.010)	0.065 (0.009)	0.066 (0.008)	0.039 (0.007)	0.025 (0.004)	0.014 (0.004)

Factors: 2 *Model order*: 1 *Variance function*: *U−shape*



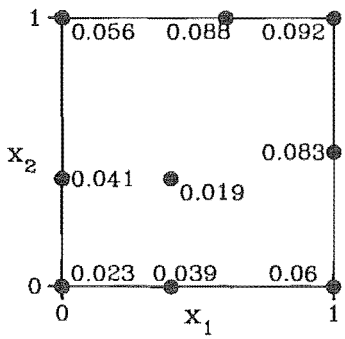
n_0 \ α	0.5	1	1.5	2	4	6
4	0.345 (0.015)	0.346 (0.012)	0.361 (0.012)	0.362 (0.012)	0.318 (0.013)	0.260 (0.012)
10	0.137 (0.010)	0.140 (0.009)	0.146 (0.009)	0.148 (0.008)	0.095 (0.006)	0.056 (0.005)
25	0.054 (0.006)	0.057 (0.006)	0.048 (0.006)	0.059 (0.005)	0.024 (0.004)	0.021 (0.003)

Factors: 2 *Model order*: 2 *Variance function*: *Flat*



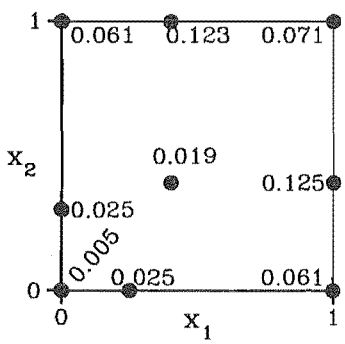
n_0 \ α	0.5	1	1.5	2	4	6
4	0.289 (0.013)	0.308 (0.012)	0.375 (0.011)	0.371 (0.011)	0.343 (0.010)	0.279 (0.010)
10	0.100 (0.010)	0.137 (0.008)	0.148 (0.007)	0.142 (0.006)	0.101 (0.005)	0.062 (0.004)
25	0.043 (0.006)	0.056 (0.005)	0.057 (0.004)	0.059 (0.004)	0.032 (0.003)	0.022 (0.002)

Factors: 2 *Model order*: 2 *Variance function*: *Medium*



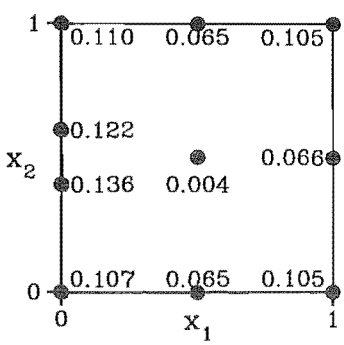
n_0 \ α	0.5	1	1.5	2	4	6
4	0.245 (0.012)	0.316 (0.012)	0.334 (0.011)	0.361 (0.012)	0.348 (0.011)	0.304 (0.011)
10	0.090 (0.010)	0.133 (0.008)	0.137 (0.008)	0.151 (0.007)	0.109 (0.006)	0.072 (0.005)
25	0.030 (0.006)	0.058 (0.006)	0.074 (0.005)	0.057 (0.005)	0.034 (0.003)	0.020 (0.003)

Factors: 2 *Model order*: 2 *Variance function*: *Steep*



n_0 \ α	0.5	1	1.5	2	4	6
4	0.233 (0.015)	0.344 (0.012)	0.374 (0.012)	0.353 (0.012)	0.334 (0.012)	0.271 (0.012)
10	0.117 (0.009)	0.141 (0.008)	0.159 (0.008)	0.157 (0.008)	0.100 (0.007)	0.069 (0.005)
25	0.039 (0.006)	0.050 (0.006)	0.063 (0.005)	0.067 (0.004)	0.024 (0.003)	0.019 (0.003)

Factors: 2 *Model order*: 2 *Variance function*: *U−shape*



n_0 \ α	0.5	1	1.5	2	4	6
4	0.261 (0.013)	0.328 (0.010)	0.355 (0.010)	0.389 (0.010)	0.370 (0.010)	0.307 (0.011)
10	0.098 (0.009)	0.149 (0.007)	0.169 (0.007)	0.170 (0.007)	0.109 (0.006)	0.074 (0.004)
25	0.030 (0.005)	0.067 (0.005)	0.080 (0.005)	0.070 (0.004)	0.041 (0.003)	0.015 (0.003)

Appendix 3: Matlab Code for 3-Phase Heuristic

```
function [x,sigs,totcost]=3phase(grid,L0,minmax,B)

% 3PHASE : Three phase heuristic solution heuristic for modified SICOED design problem

%           - stage 1: (modified) accelerated sequential design algorithm
%           - stage 2: design is 'rationalised' and made to conform to design criterion target
%           - stage 3: (modified) exchange algorithm

% Function arguments:   grid    = number of gridpoints for each factor (scalar)]
%                       L0      = design criterion target (scalar)
%                       minmax  = [x1min x1max; x2min x2max; ....] (defines design region)
%                       B       = integral of f(x)f(x)'dx

% Assumes:  - cuboidal design region
%           - equal number of grid points along each factor axis
%           - design criterion is average variance of fitted response
%           - random starting design with 2*p points

% Function calls:   f = model(x) - returns vector f(x) for candidate design point vector x
%                   obj = acostf(x) - returns scalar marginal cost (obj) for candidate design point vector x
%                   px = points(grid,minmax) - returns matrix of candidate design points

%   T.A.J. Vollebregt (1995)

% ********** Setting up problem **********

% initialise

fact=length(minmax(:,1));               % number of factors
p=length(model(zeros(fact,1)+1));       % number of parameters
gp=grid^fact;                           % total number of gridpoints

% Set up F matrix, marginal cost vector

px=points(grid,minmax);                 % define co-ordinates of gridpoints
for i=1:gp
    F(i,:)=model(px(i,:));              % f(x_i) - i^th row of F matrix
end
for i=1:gp
    obj(i)=acostf(px(i,:));             % c(x_i)v^2(x_i) - marginal cost function value at x_i
```

```
end
ub=zeros(gp,1)+inf;                     % upper bound on S (S = 1/sigma)

% ********** Phase 1: Modified sequential design algorithm **********

% Select starting design

notinvert=1;
while notinvert
    S=zeros(gp,1);                      % design problem variables - S = 1/sigma
    for i=1:2*p
        S(round(rand(1)*(gp-1))+1)=1;   % select starting design with 2*p random points
    end
    for i=1:gp
        FS(i,:)=F(i,:)*S(i);            % intermediate step, speeds calculation ( FS'*F = F'*diag(S)*F )
    end
    if rank(FS'*F)==p                   % check invertability of M for starting design
        notinvert=0;
    end
end

% Initialise Dnew (covariance matrix, M^-1), dc (design criterion value)

Dnew=inv(FS'*F);                            % covariance matrix
dc=trace(Dnew*B);                           % design criterion value

% Main loop

count=2;                            % iteration number
list=zeros(gp,1)+inf;               % list of last calculated dL(E)/dS(i)
step=1;                             % step size for each iteration
j=0;                                % point added to
going=1;
while going

    % Initialise

    Dold=Dnew;
    dcadd=zeros(1,gp);              % dcadd(i) = value of dc by adding step to S(i)

    % Find the point j to change, adjust S and Dnew

    checkedj=0;
    notfound=1;
```

```
    while notfound
        [val,j]=max(list);          % find best value in list
        if checkedj==j              % if max(list)=list(j) was updated at last iteration, choose j
            notfound=0;
        else                        % else update list(j) to see if it is really max(list)
            Dnewj=(eye(p)-((step*Dold*F(j,:)'*F(j,:))/(1+step*F(j,:)*Dold*F(j,:)')))*Dold;    % Dykstra's identity
            list(j)=(dc-trace(Dnewj*B))/obj(j);
            checkedj=j;
        end
    end
    rstep=max([step S(j)*0.1]);                                                    % (real) step taken
    S(j)=S(j)+rstep;                                                               % add step to S(j)
    Dnew=(eye(p)-((rstep*Dold*F(j,:)'*F(j,:))/(1+rstep*F(j,:)*Dold*F(j,:)')))*Dold;    % Dykstra's identity
    dc=trace(Dnew*B);                                                              % recalculate dc

    % Stopping rule

    sS=sort(S);
    if sS(gp-p+1)>3     % if p design points have been added to 4 times
        going=0;
    end
    count=count+1;
end

% ********** Phase 2: 'Rationalising' design **********

% select design points to remove from consideration

if gp>4*p                                                   % do only for suffciently large number of gridpoints
    fS=find(S);                                             % fS is list of non-zero elements of S
    for k=1:fact
        dist(k)=1.1*(minmax(k,2)-minmax(k,1))/(grid-1);     % 1.1 * distance between gridpoints along factor k axis
    end
    for i=1:length(fS)                                      % point i
        for j=i+1:length(fS)                                % point j
            if ub(fS(i))>0 & ub(fS(j))>0                    % if neither points is already excluded
                oneapart=zeros(fact,1);
                for k=1:fact                                % factor k
                    if abs(px(fS(i),k)-px(fS(j),k))<=dist(k)    % if distance is within dist(k)
                        oneapart(k)=1;
                    end
                end
                if min(oneapart)==1
                    if S(fS(i))>=S(fS(j))                   % choose larger S value
```

```
                    ub(fS(j))=0;                        % set upper bound to zero for j
                else
                    ub(fS(i))=0;                        % set upper bound to zero for i
                end
            end
        end
    end
end
for i=1:gp
    if S(i)==0
        ub(i)=0;                    % set upper bound to zero for all points not added to
    end
end

% remove gridpoints with ub(i)=0

num=0;
for i=1:gp
    if ub(i)>0
        num=num+1;
        newS(num,1)=S(i);
        newpx(num,:)=px(i,:);
        newF(num,:)=F(i,:);
        newobj(num)=obj(i);
    end
end
S=newS;
px=newpx;
F=newF;
obj=newobj;
gp=num;
FS=FS(1,:);

% scale design to meet L0, recalculate Dnew and dc

for i=1:gp
    FS(i,:)=F(i,:)*S(i);
end
Dnew=inv(FS'*F);
dc=trace(Dnew*B);
S=S*dc/L0;
Dnew=L0/dc*Dnew;
dc=L0;
```

```
% ********** Phase 2: Modified exchange algorithm **********

b(1)=obj*S;             % initialise cost (budget)
count=2;

% Initialise step parameters

step=S*0.25;            % set initial step size vector
maxstep=step;
j=0;
lastj=0;
laststep=0;             % records sign of last step
tstep=0;                % records sign of step before laststep

% Main loop

going=1;
while going

    % Initialise

    Dold=Dnew;
    dcadd=zeros(1,gp);              % dcadd(i) = dc by adding delta to S(i)

    % Set sign of movement

    if dc<L0
        maxstep=-1*abs(maxstep);
        posneg=-1;                  % to record that negative step is to be taken
    else
        maxstep=abs(maxstep);
        posneg=1;                   % to record that positive step is to be taken
    end

    % Set stepsize

    for i=1:gp
        if maxstep(i)<0
            step(i)=max([-S(i) maxstep(i)]);
        else
            step(i)=maxstep(i);
        end
        if step(i)==0
```

```
        step(i)=1e10;           % records that no step is to be allowed
    end
end

% Main loop calculating the 'gradients'

delta=posneg*0.0001*max(S);         % increment
for i=1:gp
    if step(i)<1e10
        Dnewi=(eye(p)-((delta*Dold*F(i,:)'*F(i,:))/(1+delta*F(i,:)*Dold*F(i,:)')))*Dold;  % Dykstra's identity
        dcadd(i)=trace(Dnewi*B);
    else
        dcadd(i)=inf;
    end
end

% Find the point j to change, adjust S and Dnew

lastj=j;                                                                          % record last point added to
[val,j]=max((dc-dcadd)./obj);                                                     % j is optimum point to add to
S(j)=S(j)+step(j);                                                                % add step to S(j)
Dnew=(eye(p)-((step(j)*Dold*F(j,:)'*F(j,:))/(1+step(j)*F(j,:)*Dold*F(j,:)')))*Dold;   % Dykstra's identity
dc=trace(Dnew*B);                                                                 % recalculate dc

% Change the maxstep size

if (lastj==j & laststep*step(j)<0)                          % if + then - step at same point
    maxstep(j)=maxstep(j)*0.5;
elseif (lastj==j & (laststep*step(j)>0 & tstep*laststep>0))  % if ++ or -- at same point
    maxstep(j)=maxstep(j)*2;
else
    maxstep(j)=maxstep(j)*0.99;                             % default reduction
end
tstep=laststep;
laststep=maxstep(j);

% Stopping rule

b(count)=obj*S;
if count>10
    if dc<L0
    if (max(b(max([count-10 1]):count))-min(b(max([count-10 1]):count)))/mean(b(max([count-10 1]):count))<0.005
        going=0;
    end
```

```
        end
    end
    count=count+1;
end

% reporting variables

num=0;
clear x
for i=1:gp
    if S(i)>0
        num=num+1;
        x(num,1:fact)=px(i,:);
        sigs(num)=1/S(i);
    end
end
totcost=obj*S;
```